Sugar and Lipid Metabolism Regulators in Plants IV

CROSS REFERENCE TO RELATED APPLICATIONS

[001] The present invention claims the priority benefit of U.S. Provisional Patent Application Serial No. 60/400,803 filed August 2, 2002, the entire contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

[002] This invention relates generally to nucleic acid sequences encoding proteins that are related to the presence of seed storage compounds in plants. More specifically, the present invention relates to nucleic acid sequences encoding sugar and lipid metabolism regulator proteins and the use of these sequences in transgenic plants. The invention further relates to methods of applying these novel plant polypeptides to the identification and stimulation of plant growth and/or to the increase of yield of seed storage compounds.

Background Art

[003] The study and genetic manipulation of plants has a long history that began even before the famed studies of Gregor Mendel. In perfecting this science, scientists have accomplished modification of particular traits in plants ranging from potato tubers having increased starch content to oilseed plants such as canola and sunflower having increased or altered fatty acid content. With the increased consumption and use of plant oils, the modification of seed oil content and seed oil levels has become increasingly widespread (e.g. Töpfer et al., 1995, Science 268:681-686). Manipulation of biosynthetic pathways in transgenic plants provides a number of opportunities for molecular biologists and plant biochemists to affect plant metabolism giving rise to the production of specific higher-value products. The seed oil production or composition has been altered in numerous traditional

oilseed plants such as soybean (U.S. Patent No. 5,955,650), canola (U.S. Patent No. 5,955,650), sunflower (U.S. Patent No. 6,084,164), rapeseed (Töpfer et al., 1995, Science 268:681-686), and non-traditional oil seed plants such as tobacco (Cahoon et al., 1992, Proc. Natl. Acad. Sci. USA 89:11184-11188).

[004] Plant seed oils comprise both neutral and polar lipids (See Table 1). The neutral lipids contain primarily triacylglycerol, which is the main storage lipid that accumulates in oil bodies in seeds. The polar lipids are mainly found in the various membranes of the seed cells, e.g. the endoplasmic reticulum, microsomal membranes, and the cell membrane. The neutral and polar lipids contain several common fatty acids (See Table 2) and a range of less common fatty acids. The fatty acid composition of membrane lipids is highly regulated and only a select number of fatty acids are found in membrane lipids. On the other hand, a large number of unusual fatty acids can be incorporated into the neutral storage lipids in seeds of many plant species (Van de Loo F.J. et al., 1993, Unusual Fatty Acids in Lipid Metabolism in Plants pp. 91-126, editor TS Moore Jr. CRC Press; Millar et al., 2000, Trends Plant Sci. 5:95-101).

Table 1

Plant Lipid Classes

Neutral Lipids	Triacylglycerol (TAG)
	Diacylglycerol (DAG)
	Monoacylglycerol (MAG)
Polar Lipids	Monogalactosyldiacylglycerol (MGDG)
	Digalactosyldiacylglycerol (DGDG)
	Phosphatidylglycerol (PG)
	Phosphatidylcholine (PC)
	Phosphatidylethanolamine (PE)
	Phosphatidylinositol (PI)
	Phosphatidylserine (PS)
	Sulfoquinovosyldiacylglycerol

Table 2

Common Plant Fatty Acids

16:0	Palmitic acid
16:1	Palmitoleic acid
16:3	Palmitolenic acid
18:0	Stearic acid
18:1	Oleic acid
18:2	Linoleic acid
18:3	Linolenic acid
γ-18:3	Gamma-linolenic acid*
20:0	Arachidic acid
20:1	Eicosenoic acid
22:6	Docosahexanoic acid (DHA) *
20:2	Eicosadienoic acid
20:4	Arachidonic acid (AA) *
20:5	Eicosapentaenoic acid (EPA) *
22:1	Erucic acid

[005] In Table 2, the fatty acids denoted with an asterisk do not normally occur in plant seed oils, but their production in transgenic plant seed oil is of importance in plant biotechnology.

[006] Lipids are synthesized from fatty acids, and their synthesis may be divided into two parts: the prokaryotic pathway and the eukaryotic pathway (Browse et al., 1986, Biochemical J. 235:25-31; Ohlrogge & Browse, 1995, Plant Cell 7:957-970). The prokaryotic pathway is located in plastids, the primary site of fatty acid biosynthesis. Fatty acid synthesis begins with the conversion of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (ACCase). Malonyl-CoA is converted to malonyl-acyl carrier protein (ACP) by the malonyl-CoA:ACP transacylase. The enzyme beta-keto-acyl-ACP-synthase III (KAS III) catalyzes a condensation reaction in which the acyl group from acetyl-CoA is transferred to malonyl-ACP to form 3-ketobutyryl-ACP. In a subsequent series of condensation, reduction and dehydration reactions the nascent fatty acid chain on the ACP cofactor is elongated by the step-by-step addition (condensation) of two carbon atoms donated by malonyl-ACP until a 16-carbon or 18-carbon saturated fatty acid chain is formed. The plastidial delta-9 acyl-ACP desaturase introduces the first unsaturated double bond into the fatty acid. Thioesterases cleave the fatty acids from the ACP cofactor, and free fatty acids are exported to the cytoplasm where they participate as fatty acyl-CoA esters in the eukaryotic pathway. In the eukaryotic pathway, the fatty acids are esterified by glycerol-3-phosphate acyltransferase and

lysophosphatidic acid acyltransferase to the sn-1 and sn-2 positions of glycerol-3-phosphate, respectively, to yield phosphatidic acid (PA). The PA is the precursor for other polar and neutral lipids, the latter being formed in the Kennedy pathway (Voelker, 1996, Genetic Engineering ed.:Setlow 18:111-113; Shanklin & Cahoon, 1998, Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:611-641; Frentzen, 1998, Lipids 100:161-166; Millar et al., 2000, Trends Plant Sci. 5:95-101).

[007] Storage lipids in seeds are synthesized from carbohydrate-derived precursors. Plants have a complete glycolytic pathway in the cytosol (Plaxton, 1996, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:185-214), and it has been shown that a complete pathway also exists in the plastids of rapeseeds (Kang & Rawsthorne, 1994, Plant J. 6:795-805). Sucrose is the primary source of carbon and energy, transported from the leaves into the developing seeds. During the storage phase of seeds, sucrose is converted in the cytosol to provide the metabolic precursors glucose-6-phosphate and pyruvate. These are transported into the plastids and converted into acetyl-CoA that serves as the primary precursor for the synthesis of fatty acids. Acetyl-CoA in the plastids is the central precursor for lipid biosynthesis. Acetyl-CoA can be formed in the plastids by different reactions, and the exact contribution of each reaction is still being debated (Ohlrogge & Browse, 1995, Plant Cell 7:957-970). It is accepted, however, that a large part of the acetyl-CoA is derived from glucose-6-phospate and pyruvate that are imported from the cytoplasm into the plastids. Sucrose is produced in the source organs (leaves, or anywhere that photosynthesis occurs) and is transported to the developing seeds that are also termed sink organs. In the developing seeds, the sucrose is the precursor for all the storage compounds, i.e. starch, lipids and partly the seed storage proteins. Therefore, it is clear that carbohydrate metabolism in which sucrose plays a central role is very important to the accumulation of seed storage compounds.

[008] Although lipid and fatty acid content of seed oil can be modified by the traditional methods of plant breeding, the advent of recombinant DNA technology has allowed for easier manipulation of the seed oil content of a plant, and in some cases, has allowed for the alteration of seed oils in ways that could not be accomplished by breeding alone (See, e.g., Töpfer et al. 1995, Science 268:681-686). For example, introduction of a Δ^{12}-hydroxylase nucleic acid sequence into transgenic tobacco resulted in the introduction of a novel fatty acid, ricinoleic acid, into the tobacco seed oil (Van de Loo et al., 1995, Proc. Natl. Acad. Sci USA 92:6743-6747). Tobacco plants have also been engineered to produce

low levels of petroselinic acid by the introduction and expression of an acyl-ACP desaturase from coriander (Cahoon et al., 1992, Proc. Natl. Acad. Sci USA 89:11184-11188).

[009] The modification of seed oil content in plants has significant medical, nutritional, and economic ramifications. With regard to the medical ramifications, the long chain fatty acids (C18 and longer) found in many seed oils have been linked to reductions in hypercholesterolemia and other clinical disorders related to coronary heart disease (Brenner, 1976, Adv. Exp. Med. Biol. 83:85-101). Therefore, consumption of a plant having increased levels of these types of fatty acids may reduce the risk of heart disease. Enhanced levels of seed oil content also increase large-scale production of seed oils and thereby reduce the cost of these oils.

[010] In order to increase or alter the levels of compounds such as seed oils in plants, nucleic acid sequences and proteins regulating lipid and fatty acid metabolism must be identified. As mentioned earlier, several desaturase nucleic acids such as the Δ^6-desaturase nucleic acid, Δ^{12}-desaturase nucleic acid and acyl-ACP desaturase nucleic acid have been cloned and demonstrated to encode enzymes required for fatty acid synthesis in various plant species. Oleosin nucleic acid sequences from such different species as *Brassica*, soybean, carrot, pine, and *Arabidopsis thaliana* have also been cloned and determined to encode proteins associated with the phospholipid monolayer membrane of oil bodies in those plants.

[011] It has also been determined that two phytohormones, gibberellic acid (GA) and absisic acid (ABA), are involved in overall regulatory processes in seed development (e.g. Ritchie & Gilroy, 1998, Plant Physiol. 116:765-776; Arenas-Huertero et al., 2000, Genes Dev. 14:2085-2096). Both the GA and ABA pathways are affected by okadaic acid, a protein phosphatase inhibitor (Kuo et al., 1996, Plant Cell. 8:259-269). The regulation of protein phosphorylation by kinases and phosphatases is accepted as a universal mechanism of cellular control (Cohen, 1992, Trends Biochem. Sci. 17:408-413). Likewise, the plant hormones ethylene (e.g. Zhou et al., 1998, Proc. Natl. Acad. Sci. USA 95:10294-10299; Beaudoin et al., 2000, Plant Cell 2000:1103-1115), and auxin (e.g. Colon-Carmona et al., 2000, Plant Physiol. 124:1728-1738) are involved in controlling plant development as well.

[012] Although several compounds are known that generally affect plant and seed development, there is a clear need to specifically identify factors that are more specific for the developmental regulation of storage compound accumulation and to identify genes which have the capacity to confer altered or increased oil production to its host plant and to other plant species. This invention discloses a large number of nucleic acid sequences from

Arabidopsis thaliana, Brassica napus, and the moss *Physcomitrella patens*. These nucleic acid sequences can be used to alter or increase the levels of seed storage compounds such as proteins, sugars and oils, in plants, including transgenic plants, such as rapeseed, canola, linseed, soybean, sunflower maize, oat, rye, barley, wheat, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor and peanut, which are oilseed plants containing high amounts of lipid compounds.

SUMMARY OF THE INVENTION

[013] The present invention provides novel isolated nucleic acid and amino acid sequences associated with the metabolism of seed storage compounds in plants.

[014] The present invention also provides an isolated nucleic acid from Arabidopsis, Brassica, and *Physcomitrella patens* encoding a Lipid Metabolism Protein (LMP), or a portion thereof. These sequences may be used to modify or increase lipids and fatty acids, cofactors and enzymes in microorganisms and plants.

[015] Arabidopsis plants are known to produce considerable amounts of fatty acids such as linoleic and linolenic acid (See, e.g., Table 2) and for their close similarity in many aspects (gene homology, etc.) to the oil crop plant Brassica. Therefore, nucleic acid molecules originating from a plant like *Arabidopsis thaliana* and *Brassica napus* are especially suited to modify the lipid and fatty acid metabolism in a host, especially in microorganisms and plants. Furthermore, nucleic acids from the plants *Arabidopsis thaliana* and *Brassica napus* can be used to identify those DNA sequences and enzymes in other species which are useful to modify the biosynthesis of precursor molecules of fatty acids in the respective organisms.

[016] The present invention further provides an isolated nucleic acid comprising a fragment of at least 15 nucleotides of a nucleic acid from a plant (*Arabidopsis thaliana, Brassica napus, or Physcomitrella patens*) encoding a Lipid Metabolism Protein (LMP), or a portion thereof.

[017] Also provided by the present invention are polypeptides encoded by the nucleic acids, heterologous polypeptides comprising polypeptides encoded by the nucleic acids, and antibodies to those polypeptides.

[018] Additionally, the present invention relates to and provides the use of LMP nucleic acids in the production of transgenic plants having a modified level of a seed storage compound. A method of producing a transgenic plant with a modified level of a seed storage compound includes the steps of transforming a plant cell with an expression vector comprising a LMP nucleic acid, and generating a plant with a modified level of the seed

storage compound from the plant cell. In a preferred embodiment, the plant is an oil producing species selected from the group consisting of rapeseed, canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, and peanut, for example.

[019] According to the present invention, the compositions and methods described herein can be used to increase or decrease the level of an LMP in a transgenic plant comprising increasing or decreasing the expression of the LMP nucleic acid in the plant. Increased or decreased expression of the LMP nucleic acid can be achieved through *in vivo* mutagenesis of the LMP nucleic acid. The present invention can also be used to increase or decrease the level of a lipid in a seed oil, to increase or decrease the level of a fatty acid in a seed oil, or to increase or decrease the level of a starch in a seed or plant.

[020] Also included herein is a seed produced by a transgenic plant transformed by a LMP DNA sequence, wherein the seed contains the LMP DNA sequence and wherein the plant is true breeding for a modified level of a seed storage compound. The present invention additionally includes a seed oil produced by the aforementioned seed.

[021] Further provided by the present invention are vectors comprising the nucleic acids, host cells containing the vectors, and descendent plant materials produced by transforming a plant cell with the nucleic acids and/or vectors.

[022] According to the present invention, the compounds, compositions, and methods described herein can be used to increase or decrease the level of a lipid in a seed oil, or to increase or decrease the level of a fatty acid in a seed oil, or to increase or decrease the level of a starch or other carbohydrate in a seed or plant. A method of producing a higher or lower than normal or typical level of storage compound in a transgenic plant, comprises expressing a LMP nucleic acid from *Arabidopsis thaliana, Brassica napus*, and *Physcomitrella patens* in the transgenic plant, wherein the transgenic plant is *Arabidopsis thaliana* and *Brassica napus*, or a species different from *Arabidopsis thaliana* and *Brassica napus*. Also included herein are compositions and methods of the modification of the efficiency of production of a seed storage compound. As used herein, the phrase "*Arabidopsis thaliana* and *Brassica napus*" means *Arabidopsis thaliana* and/or *Brassica napus*.

[023] Accordingly, the present invention provides novel isolated LMP nucleic acids and isolated LMP amino acid sequences from *Arabidopsis thaliana, Brassica napus*, and *Physcomitrella patens*, as well as active fragments, analogs and orthologs thereof.

[024] The present invention also provides transgenic plants having modified levels of seed storage compounds, and in particular, modified levels of a lipid, a fatty acid, or a sugar.

[025] The polynucleotides and polypeptides of the present invention, including agonists and/or fragments thereof, also have uses that include modulating plant growth, and potentially plant yield, preferably increasing plant growth under adverse conditions (drought, cold, light, UV). In addition, antagonists of the present invention may have uses that include modulating plant growth and/or yield, preferably through increasing plant growth and yield. In yet another embodiment, overexpression of the polypeptides of the present invention using a constitutive promoter (e.g., 35S or other promoters) may be useful for increasing plant yield under stress conditions (drought, light, cold, UV) by modulating light utilization efficiency.

[026] The present invention also provides methods for producing such aforementioned transgenic plants. In another embodiment, the present invention provides seeds and seed oils from such aforementioned transgenic plants.

[027] These and other embodiments, features, and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and the appended claims.

DETAILED DESCRIPTION OF THE INVENTION

[028] The present invention may be understood more readily by reference to the following detailed description of the preferred embodiments of the invention and the Examples included therein.

[029] Before the present compounds, compositions, and methods are disclosed and described, it is to be understood that this invention is not limited to specific nucleic acids, specific polypeptides, specific cell types, specific host cells, specific conditions, or specific methods, etc., as such may, of course, vary, and the numerous modifications and variations therein will be apparent to those skilled in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting. As used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

[030] In accordance with the purpose(s) of this invention, as embodied and broadly described herein, this invention, in one aspect, provides an isolated nucleic acid from a plant

(*Arabidopsis thaliana, Brassica napus*, and *Physcomitrella patens*) encoding a Lipid Metabolism Protein (LMP), or a portion thereof. As used herein, the phrase "*Arabidopsis thaliana, Brassica napus*, and *Physcomitrella patens*" means *Arabidopsis thaliana* and/or *Brassica napus* and/or *Physcomitrella patens*.

[031] One aspect of the invention pertains to isolated nucleic acid molecules that encode LMP polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of an LMP-encoding nucleic acid (e.g., LMP DNA). As used herein, the terms "nucleic acid molecule" and "polynucleotide sequence" are used interchangeably and are intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of a gene: at least about 1000 nucleotides of sequence upstream from the 5' end of the coding region and at least about 200 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is substantially separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is substantially free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated LMP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., an *Arabidopsis thaliana* or *Brassica napus* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors, or other chemicals when chemically synthesized.

[032] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a polynucleotide sequence of Appendix A (i.e. the polynucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID

NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, an *Arabidopsis thaliana*, *Brassica napus*, or *Physcomitrella patens* LMP cDNA can be isolated from an *Arabidopsis thaliana*, *Brassica napus*, or *Physcomitrella patens* library using all or portion of one of the polynucleotide sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Moreover, a nucleic acid molecule encompassing all or a portion of one of the polynucleotide sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from plant cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al., 1979, Biochemistry 18:5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the polynucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to a LMP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[033] In a preferred embodiment, an isolated nucleic acid of the invention comprises one of the polynucleotide sequences shown in Appendix A (i.e. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID

NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81). These polynucleotides of Appendix A correspond to the *Arabidopsis thaliana, Brassica napus*, and *Physcomitrella patens* LMP cDNAs of the invention. These cDNAs comprise sequences encoding LMPs (i.e., the "coding region" or open reading frame (ORF)), as well as 5' untranslated sequences and 3' untranslated sequences. Alternatively, the nucleic acid molecules can comprise only the coding region of any of the polynucleotide sequences described herein or can contain whole genomic fragments isolated from genomic DNA.

[034] For the purposes of this application, it will be understood that each of the polynucleotide sequences set forth in Appendix A has an identifying entry number (e.g., Pk123). Each of these sequences may generally comprise three parts: a 5' upstream region, a coding region, and a downstream region. The particular polynucleotide sequences shown in Appendix A represent the coding region or open reading frame, and the putative functions of the encoded polypeptides are indicated in Table 3.

Table 3

Putative LMP Functions

Sequence code	Function	SEQ ID NO:
Pk123	Gibberellin-regulated protein GASA3 precursor	1
Pk197	Tyrosine aminotransferase	3
Pk136	D-hydroxy-fatty acid dehydrogenase	5
Pk156	Serine protease	7
Pk159	Nonspecific lipid-transfer protein	9
Pk179	Signal transduction protein	11
Pk202	Lipid transfer - like protein	13
Pk206	bZIP transcription factor	15
Pk207	Acyl-CoA dehydrogenase	17
Pk209	Pyruvate kinase	19
Pk215	Phosphatidylglycerotransferase	21
Pk239	Digalactosyldiacylglycerol synthase	23
Pk240	Phosphatidate cytidyltransferase	25
Pk241	AT Psbs protein	27
Pk242	Omega-6 fatty acid desaturase, endoplasmic reticulum (FAD2)	29
Bn011	Gibberellin 3-beta hydroxylase with +4 G	31
Bn077	Zinc finger DNA binding protein	33

Jb001	Gibberellin 20-oxidase	35
Jb002	Seed maturation protein	37
Jb003	Beta-VPE Vacuolar Processing Enzym	39
Jb005	Very-long-chain fatty acid condensing enzyme CUT1	41
Jb007	Glucokinase	43
Jb009	Glutathione S-transferase TSI-1	45
Jb013	ABA-regulated gene	47
Jb017	Cysteine proteinase	51
Jb024	Pectinesterase-like protein	53
Jb027	Signal transduction protein	55
OO-1	Aldose reductase-like protein	57
OO-2	Dormancy related protein	59
OO-3	HSP associated protein like	61
OO-4	Poly (ADP-ribose) polymerase	63
OO-5	Transitional endoplasmic reticulum ATPase	65
OO-6	Beta coat like protein	67
OO-8	Protein disulfide-isomerase	69
OO-9	Signal transduction protein/Apoptosis inhibitor	71
OO-10	Annexin	73
OO-11	Putative oxidoreductase	75
OO-12	Long chain alc dehydrogenase/ oxidoreductase	77
pp82	Transcription factor	79
Pk225	Amino-cyclopropane-carboxylic acid oxidase	81

Table 4

Grouping of LMPs based on Functional protein domains

Functional category	SEQ ID:	SEQ Code:	Functional domain	Domain position
DNA-binding proteins	1	Pk123	Zinc finger	66-86 29-71
	15	Pk206	bZIP transcription factor (PFAM) Leucine zipper	144-197 179-209
	27	Pk241	DNA-binding domain Histone H5 signature	207-221 57-71
	33	Bn077	Zinc finger (BRCT; PARP) Ethylene responsive element binding protein	64-104 79-99
	63	OO-4	Zinc finger Leucine zipper	760-805 114-117
	73	OO-10	Zinc finger Yeast DNA-binding domain	220-230 207-217

	79	pp82	Myb DNA-binding domain	19-119
Kinases	43	Jb007	Glucokinase	173-206
	45	Jb009	Deoxynucleoside kinase	99-139
	19	Pk209	Pyruvate kinase (PFAM)	1-326
	61	OO-3	Galactokinase	285-296
Signal Transduction	67	OO-6	Wnt-1 domain WSC domain	607-655 527-548
	71	OO-9	BIR repeat (inhibitor of apoptosis) Wnt-1 domain	47-85 43-91
	41	Jb005	Wnt-1 domain	23-71
	47	Jb013	Wnt-1 domain	23-91
	55	Jb027	Emp24/gp25L intracellular vesicle trafficking Wnt-1 domain	2-204 135-183
	11	Pk179	Wnt-1 domain PDZ domain (Wnt signalling)	279-327 205-299
	3	Pk197	Wnt-1 domain	300-348
Proteases	7	Pk156	Serine protease Prolyl aminopeptidase	171-191 128-139
	37	Jb002	Peptidase family M23/M37	404-444
	39	Jb003	Cysteine protease Peptidase C13 (PFAM)	52-76 10-367
	51	Jb017	Cysteine protease C1 Peptidase C1 (PFAM)	163-178 145-361
	65	OO-5	Peptidase family M41 AAA ATPase molecular chaperone (PFAM)	343-387 620-664 243-427
Lipid metabolism	5	Pk136	D-Hydroxy-fatty acid dehydrogenase	94-143
	9	Pk159	Lipid Transfer Protein LTP (PFAM)	29-117
	13	Pk202	Lipid Transfer Protein LTP (PFAM)	38-103
	17	Pk207	Acyl-CoA dehydrogenase Iron-containing alcohol dehydrogenase	2-44 97-112
	21	Pk215	CDP-alcohol phosphatidyltransferase (PFAM)	172-309
	23	Pk239	Glycosyl (galactosyl) transferase (PFAM)	572-674
	25	Pk240	Phosphatidate cytidyltransferase	343-370
	29	Pk242	Fatty acid desaturase (PFAM)	32-376
Oxido-reductases	31	Bn011	Iron Ascorbate oxidoreductase (PFAM)	43-343
	35	Jb001	Respiratory chain NADH dehydrogenase Iron Ascorbate oxidoreductase (PFAM)	95-123 54-369
	53	Jb024	Multicopper oxidase Copper-oxidase (PFAM)	216-247 123-145 154-306
	57	OO-1	Aldo/keto reductase family (PFAM)	18-294

	59	OO-2	Alcohol dehydrogenase (PFAM)	38-228
	69	OO-8	Thioredoxin (PFAM)	22-250
	75	OO-11	Alcohol dehydrogenase (PFAM)	50-234
	77	OO-12	Zinc alcohol dehydrogenase(PFAM)	20-329
	81	Pk225	Iron Ascorbate oxidoreductase (PFAM)	3-297

[035] In another preferred embodiment, an isolated nucleic acid molecule of the present invention encodes a polypeptide that is able to participate in the metabolism of seed storage compounds such as lipids, starch, and seed storage proteins, and that contains a DNA-binding (or transcription factor) domain, a protein kinase domain, a signal transduction domain, a protease domain, a lipid metabolism domain, or an oxidoreductase domain. Examples of isolated nucleic acids that encode LMPs containing such domains can be found in Table 4. Examples of nucleic acids encoding LMPs containing a DNA-binding domain include those shown in SEQ ID NO:1, SEQ ID NO:15, SEQ ID NO:27, SEQ ID NO:33, SEQ ID NO:63, SEQ ID NO:73, and SEQ ID NO:79. Examples of nucleic acids encoding LMPs containing a protein kinase domain include those shown in SEQ ID NO:19, SEQ ID NO:43, SEQ ID NO:45, and SEQ ID NO:61. Examples of nucleic acids encoding LMPs containing a signal transduction domain include those shown in SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:41, SEQ ID NO:47, SEQ ID NO:55, SEQ ID NO:67, and SEQ ID NO:71. Examples of nucleic acids encoding LMPs containing a protease domain include those shown in SEQ ID NO:7, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:51, and SEQ ID NO:65. Examples of nucleic acids encoding LMPs containing a lipid metabolism domain include those shown in SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:13, SEQ ID NO:17, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:29. Examples of nucleic acids encoding LMPs containing a oxidoreductase domain include those shown in SEQ ID NO:31, SEQ ID NO:35, SEQ ID NO:53, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:69, SEQ ID NO:75, SEQ ID NO:77, and SEQ ID NO:81.

[036] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule, which is a complement of one of the polynucleotide sequences shown in Appendix A (i.e. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51,

SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81), or a portion thereof. A nucleic acid molecule which is complementary to one of the polynucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the polynucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

[037] In another preferred embodiment, an isolated nucleic acid of the invention comprises a polynucleotide sequence encoding a polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, SEQ ID NO:70, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:78, SEQ ID NO:80, or SEQ ID NO:82.

[038] In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a polynucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, or more homologous to a polynucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a polynucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the polynucleotide sequences shown in Appendix A, or a portion thereof. These stringent conditions include washing with a solution having a salt concentration of about 0.02 M at pH 7 and about 60°C. In another embodiment, the stringent conditions comprise an initial hybridization in a 6X sodium chloride/sodium citrate (6X SSC) solution at 65°C.

[039] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a LMP. The polynucleotide sequences determined from the cloning of the LMP genes from *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* allows for the

generation of probes and primers designed for use in identifying and/or cloning LMP homologues in other cell types and organisms, as well as LMP homologues from other plants or related species. Therefore this invention also provides compounds comprising the nucleic acids disclosed herein, or fragments thereof. These compounds include the nucleic acids attached to a moiety. These moieties include, but are not limited to, detection moieties, hybridization moieties, purification moieties, delivery moieties, reaction moieties, binding moieties, and the like. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50, or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a polynucleotide sequence of Appendix A can be used in PCR reactions to clone LMP homologues. Probes based on the LMP nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a genomic marker test kit for identifying cells which express a LMP, such as by measuring a level of a LMP-encoding nucleic acid in a sample of cells, e.g., detecting LMP mRNA levels or determining whether a genomic LMP gene has been mutated or deleted.

[040] In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid encoded by a sequence of Appendix A such that the protein or portion thereof maintains the same or a similar function as the wild-type protein. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent amino acid residues to an amino acid sequence such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the production of seed storage compounds in plants, construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes. As used herein, an "equivalent" amino acid residue is, for example., an amino acid residue which has a similar side chain as a particular amino acid residue that is encoded by a polynucleotide sequence of Appendix A. Regulatory proteins, such as DNA binding proteins, transcription factors, kinases, phosphatases, or protein members of metabolic pathways such as the lipid, starch and protein

biosynthetic pathways, or membrane transport systems, may play a role in the biosynthesis of seed storage compounds. Examples of such activities are described herein (see putative annotations in Table 3). Examples of LMP-encoding nucleic acid sequences are set forth in Appendix A.

[041] As altered or increased sugar and/or fatty acid production is a general trait wished to be inherited into a wide variety of plants like maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower and tagetes, solanaceous plants like potato, tobacco, eggplant, and tomato, Vicia species, pea, alfalfa, bushy plants (coffee, cacao, tea), Salix species, trees (oil palm, coconut), perennial grasses, and forage crops, these crop plants are also preferred target plants for genetic engineering as one further embodiment of the present invention. As used herein, a "forage crop" includes, but is not limited to, Wheatgrass, Canarygrass, Bromegrass, Wildrye Grass, Bluegrass, Orchardgrass, Alfalfa, Salfoin, Birdsfoot Trefoil, Alsike Clover, Red Clover, and Sweet Clover.

[042] Portions of proteins encoded by the LMP nucleic acid molecules of the invention are preferably biologically active portions of one of the LMPs. As used herein, the term "biologically active portion of a LMP" is intended to include a portion, e.g., a domain/motif, of a LMP that participates in the metabolism of compounds necessary for the biosynthesis of seed storage lipids, or the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes, or has an activity as set forth in Table 3. To determine whether a LMP or a biologically active portion thereof can participate in the metabolism of compounds necessary for the production of seed storage compounds and cellular membranes, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, and as described in Example 14 of the Exemplification.

[043] Biologically active portions of a LMP include peptides comprising amino acid sequences derived from the amino acid sequence of a LMP (e.g., an amino acid sequence encoded by a nucleic acid sequence of Appendix A (i.e. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID

NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81) or the amino acid sequence of a protein homologous to an LMP, which include fewer amino acids than a full length LMP or the full length protein which is homologous to an LMP) and exhibit at least one activity of an LMP. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100, or more amino acids in length) comprise a domain or motif with at least one activity of a LMP. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of a LMP include one or more selected domains/motifs or portions thereof having biological activity.

[044] Additional nucleic acid fragments encoding biologically active portions of a LMP can be prepared by isolating a portion of one of the sequences, expressing the encoded portion of the LMP or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the LMP or peptide.

[045] The invention further encompasses nucleic acid molecules that differ from one of the polynucleotide sequences shown in Appendix A (i.e. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81), and portions thereof) due to degeneracy of the genetic code and thus encode the same LMP as that encoded by the polynucleotide sequences shown in Appendix A. In a further embodiment, the nucleic acid molecule of the invention encodes a full length protein which is substantially homologous to an amino acid sequence shown in Appendix A. In one embodiment, the full-length nucleic acid or protein or fragment of the nucleic acid or protein is from *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens*.

[046] In addition to the *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* LMP polynucleotide sequences described herein, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LMPs may exist within a population (e.g., the *Arabidopsis thaliana*, and *Brassica napus*, and *Physcomitrella patens* population). Such genetic polymorphism in the

LMP gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a LMP, preferably an *Arabidopsis thaliana*, *Brassica napus*, or *Physcomitrella patens* LMP. Such natural variations can typically result in 1-40% variance in the nucleotide sequence of the LMP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in LMP that are the result of natural variation and that do not alter the functional activity of LMPs are intended to be within the scope of the invention.

[047] Nucleic acid molecules corresponding to natural variants and non-*Arabidopsis thaliana* and *Brassica napus* orthologs of the *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* LMP cDNA of the invention can be isolated based on their homology to *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* LMP nucleic acid disclosed herein using the *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. As used herein, the term "orthologs" refers to two nucleic acids from different species, but that have evolved from a common ancestral gene by speciation. Normally, orthologs encode proteins having the same or similar functions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a polynucleotide sequence shown in Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250, or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75%, or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. In another embodiment, the stringent conditions comprise an initial hybridization in a 6X sodium chloride/sodium citrate (6X SSC) solution at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a polynucleotide sequence of Appendix A (i.e. SEQ

ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a polynucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *Arabidopsis thaliana*, *Brassica napus*, or *Physcomitrella patens* LMP.

[048] In addition to naturally-occurring variants of the LMP sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a polynucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded LMP, without altering the functional ability of the LMP. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a polynucleotide sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the LMPs (Appendix A) without altering the activity of said LMP, whereas an "essential" amino acid residue is required for LMP activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having LMP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering LMP activity.

[049] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LMPs that contain changes in amino acid residues that are not essential for LMP activity. Such LMPs differ in amino acid sequence from a sequence yet retain at least one of the LMP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence encoded by a nucleic acid of Appendix A and is capable of participation in the metabolism of compounds necessary for the production of seed storage compounds in *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens*, or cellular membranes, or has one or more activities set forth in Table 3. Preferably, the protein encoded by the nucleic acid molecule is at least

about 50-60% homologous to one of the sequences encoded by a nucleic acid of Appendix A (i.e. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, or SEQ ID NO:81), more preferably at least about 60-70% homologous to one of the sequences encoded by a nucleic acid of Appendix A, even more preferably at least about 70-80%, 80-90%, or 90-95% homologous to one of the sequences encoded by a nucleic acid of Appendix A, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences encoded by a nucleic acid of Appendix A.

[050] To determine the percent homology of two amino acid sequences (e.g., one of the sequences encoded by a nucleic acid of Appendix A and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences encoded by a nucleic acid of Appendix A) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence encoded by a nucleic acid of Appendix A), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = numbers of identical positions/total numbers of positions x 100).

[051] An isolated nucleic acid molecule encoding a LMP homologous to a protein sequence encoded by a nucleic acid of Appendix A can be created by introducing one or more nucleotide substitutions, additions, or deletions into a polynucleotide sequence of Appendix A such that one or more amino acid substitutions, additions, or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one

in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in a LMP is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a LMP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for a LMP activity described herein to identify mutants that retain LMP activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Examples 13-14 of the Exemplification).

[052] LMPs are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described herein), and the LMP is expressed in the host cell. The LMP can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, a LMP or peptide thereof can be synthesized chemically using standard peptide synthesis techniques. Moreover, native LMP can be isolated from cells, for example using an anti-LMP antibody, which can be produced by standard techniques utilizing a LMP or fragment thereof of this invention.

[053] The invention also provides LMP chimeric or fusion proteins. As used herein, a LMP "chimeric protein" or "fusion protein" comprises a LMP polypeptide operatively linked to a non-LMP polypeptide. An "LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a LMP, whereas a "non-LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the LMP, e.g., a protein which is different from the LMP and which is derived from the same or a different organism. As used herein with respect to the fusion protein, the term "operatively linked" is intended to indicate that the LMP polypeptide and the non-LMP polypeptide are fused to each other so that both sequences fulfill the

proposed function attributed to the sequence used. The non-LMP polypeptide can be fused to the N-terminus or C-terminus of the LMP polypeptide. For example, in one embodiment, the fusion protein is a GST-LMP (glutathione S-transferase) fusion protein in which the LMP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant LMPs. In another embodiment, the fusion protein is a LMP containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a LMP can be increased through use of a heterologous signal sequence.

[054] Preferably, a LMP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (See, for example, Current Protocols in Molecular Biology, eds. Ausubel et al., John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An LMP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LMP.

[055] In addition to the nucleic acid molecules encoding LMPs described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire LMP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a LMP. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of Pk121 comprises

nucleotides 1 to 786). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding LMP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

[056] Given the coding strand sequences encoding LMP disclosed herein (e.g., the polynucleotide sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LMP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of LMP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LMP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. An antisense or sense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylamino-methyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydro-uracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methyl-cytosine, N-6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyl-uracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diamino-purine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted

nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[057] In another variation of the antisense technology, a double-strand interfering RNA construct can be used to cause a down-regulation of the LMP mRNA level and LMP activity in transgenic plants. This requires transforming the plants with a chimeric construct containing a portion of the LMP sequence in the sense orientation fused to the antisense sequence of the same portion of the LMP sequence. A DNA linker region of variable length can be used to separate the sense and antisense fragments of LMP sequences in the construct.

[058] The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a LMP to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic including plant promoters are preferred.

[059] In yet another embodiment, the antisense nucleic acid molecule of the invention is an anomeric nucleic acid molecule. An anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual units, the strands run parallel to each other (Gaultier et al., 1987, Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methyl-ribonucleotide (Inoue et al., 1987, Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

[060] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff & Gerlach, 1988, Nature 334:585-591)) can be used to catalytically cleave LMP

mRNA transcripts to thereby inhibit translation of LMP mRNA. A ribozyme having specificity for an LMP-encoding nucleic acid can be designed based upon the nucleotide sequence of an LMP cDNA disclosed herein (e.g., Pk123 in Appendix A) or on the basis of a heterologous sequence to be isolated according to methods taught in this invention. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a LMP-encoding mRNA (See, e.g., U.S. Patent Nos. 4,987,071 and 5,116,742 to Cech et al.). Alternatively, LMP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (See, e.g., Bartel, D. & Szostak J.W. 1993, Science 261:1411-1418).

[061] Alternatively, LMP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a LMP nucleotide sequence (e.g., a LMP promoter and/or enhancers) to form triple helical structures that prevent transcription of a LMP gene in target cells (See generally, Helene C., 1991, Anticancer Drug Des. 6:569-84; Helene C. et al., 1992, Ann. N.Y. Acad. Sci. 660:27-36; and Maher, L.J., 1992, Bioassays 14:807-15).

[062] Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a LMP (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[063] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. As used herein with respect to a recombinant expression vector, "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence and both sequences are fused to each other so that each fulfills its proposed function (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers, and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) and Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnolgy, CRC Press, Boca Raton, Florida, eds.: Glick & Thompson, Chapter 7, 89-108 including the references therein. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells or under certain conditions. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., LMPs, mutant forms of LMPs, fusion proteins, etc.).

[064] The recombinant expression vectors of the invention can be designed for expression of LMPs in prokaryotic or eukaryotic cells. For example, LMP genes can be expressed in bacterial cells, insect cells (using baculovirus expression vectors), yeast and other fungal cells (See Romanos M.A. et al., 1992, Foreign gene expression in yeast: a review, Yeast 8:423-488; van den Hondel, C.A.M.J.J. et al., 1991, Heterologous gene expression in filamentous fungi, in: More Gene Manipulations in Fungi, Bennet & Lasure, eds., p. 396-428:Academic Press: an Diego; and van den Hondel & Punt, 1991, Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae (Falciatore et al., 1999, Marine Biotechnology 1:239-251), ciliates of the types: Holotrichia, Peritrichia, Spirotrichia, Suctoria, Tetrahymena, Paramecium, Colpidium, Glaucoma,

Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella, and Stylonychia, especially of the genus Stylonychia lemnae with vectors following a transformation method as described in WO 98/01572, and multicellular plant cells (See Schmidt & Willmitzer, 1988, High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* leaf and cotyledon plants, Plant Cell Rep.:583-586; Plant Molecular Biology and Biotechnology, C Press, Boca Raton, Florida, chapter 6/7, S.71-119 (1993); White, Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung and Wu, Academic Press 1993, 128-43; Potrykus, 1991, Annu. Rev. Plant Physiol. Plant Mol. Biol. 42:205-225 (and references cited therein)), or mammalian cells. Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA 1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[065] Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve one or more of the following purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

[066] Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, 1988, Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA), and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the LMP is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant LMP unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

[067] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., 1988, Gene 69:301-315) and pET 11d (Studier et al., 1990, Gene Expression Technology:Methods in Enzymology 185, Academic Press, San Diego, California 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

[068] One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman S., 1990, Gene Expression Technology:Methods in Enzymology 185:119-128, Academic Press, San Diego, California). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression (Wada et al., 1992, Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[069] In another embodiment, the LMP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al., 1987, Embo J. 6:229-234), pMFa (Kurjan & Herskowitz, 1982, Cell 30:933-943), pJRY88 (Schultz et al., 1987, Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel & Punt, 1991, "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy et al., eds., p. 1-28, Cambridge University Press: Cambridge.

[070] Alternatively, the LMPs of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., 1983, Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow & Summers, 1989, Virology 170:31-39).

[071] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987, Nature 329:840) and pMT2PC (Kaufman et al., 1987, EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control

functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus, and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells, see chapters 16 and 17 of Sambrook, Fritsh and Maniatis, Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[072] In another embodiment, the LMPs of the invention may be expressed in unicellular plant cells (such as algae, see Falciatore et al. (1999, Marine Biotechnology 1:239-251 and references therein) and plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, Kemper, Schell and Masterson (1992, "New plant binary vectors with selectable markers located proximal to the left border", Plant Mol. Biol. 20:1195-1197) and Bevan (1984, "Binary *Agrobacterium* vectors for plant transformation, Nucleic Acids Res. 12:8711-8721; Vectors for Gene Transfer in Higher Plants; in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung und R. Wu, Academic Press, 1993, S. 15-38).

[073] A plant expression cassette preferably contains regulatory sequences capable to drive gene expression in plant cells and which are operatively linked so that each sequence can fulfil its function such as termination of transcription, including polyadenylation signals. Preferred polyadenylation signals are those originating from *Agrobacterium tumefaciens* t-DNA such as the gene 3 known as octopine synthase of the Ti-plasmid pTiACH5 (Gielen et al. 1984, EMBO J. 3:835) or functional equivalents thereof but also all other terminators functionally active in plants are suitable.

[074] As plant gene expression is very often not limited on transcriptional levels a plant expression cassette preferably contains other operatively linked sequences like translational enhancers such as the overdrive-sequence containing the 5'-untranslated leader sequence from tobacco mosaic virus enhancing the protein per RNA ratio (Gallie et al. 1987, Nucleic Acids Res. 15:8693-8711).

[075] Plant gene expression has to be operatively linked to an appropriate promoter conferring gene expression in a timely, cell or tissue specific manner. Preferred are promoters driving constitutive expression (Benfey et al. 1989, EMBO J. 8:2195-2202) like those derived from plant viruses like the 35S CAMV (Franck et al. 1980, Cell 21:285-294), the 19S CaMV (see also US 5,352,605 and WO 84/02913) or plant promoters like those from Rubisco small subunit described in US 4,962,028. Even more preferred are seed-specific promoters driving expression of LMP proteins during all or selected stages of seed development. Seed-specific

plant promoters are known to those of ordinary skill in the art and are identified and characterized using seed-specific mRNA libraries and expression profiling techniques. Seed-specific promoters include the napin-gene promoter from rapeseed (US 5,608,152), the USP-promoter from *Vicia faba* (Baeumlein et al. 1991, Mol. Gen. Genetics 225:459-67), the oleosin-promoter from Arabidopsis (WO 98/45461), the phaseolin-promoter from *Phaseolus vulgaris* (US 5,504,200), the Bce4-promoter from *Brassica* (WO9113980) or the legumin B4 promoter (LeB4; Baeumlein et al. 1992, Plant J. 2:233-239) as well as promoters conferring seed specific expression in monocot plants like maize, barley, wheat, rye, rice etc. Suitable promoters to note are the lpt2 or lpt1-gene promoter from barley (WO 95/15389 and WO 95/23230) or those described in WO 99/16890 (promoters from the barley hordein-gene, the rice glutelin gene, the rice oryzin gene, the rice prolamin gene, the wheat gliadin gene, wheat glutelin gene, the maize zein gene, the oat glutelin gene, the Sorghum kasirin-gene, and the rye secalin gene).

[076] Plant gene expression can also be facilitated via an inducible promoter (for review see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:89-108). Chemically inducible promoters are especially suitable if gene expression is desired in a time specific manner. Examples for such promoters are a salicylic acid inducible promoter (WO 95/19443), a tetracycline inducible promoter (Gatz et al. 1992, Plant J. 2:397-404) and an ethanol inducible promoter (WO 93/21334).

[077] Promoters responding to biotic or abiotic stress conditions are also suitable promoters such as the pathogen inducible PRP1-gene promoter (Ward et al., 1993, Plant. Mol. Biol. 22:361-366), the heat inducible hsp80-promoter from tomato (US 5,187,267), cold inducible alpha-amylase promoter from potato (WO 96/12814) or the wound-inducible pinII-promoter (EP 375091).

[078] Other preferred sequences for use in plant gene expression cassettes are targeting-sequences necessary to direct the gene-product in its appropriate cell compartment (for review see Kermode 1996, Crit. Rev. Plant Sci. 15:285-423 and references cited therein) such as the vacuole, the nucleus, all types of plastids like amyloplasts, chloroplasts, chromoplasts, the extracellular space, mitochondria, the endoplasmic reticulum, oil bodies, peroxisomes and other compartments of plant cells. Also especially suited are promoters that confer plastid-specific gene expression, as plastids are the compartment where precursors and some end products of lipid biosynthesis are synthesized. Suitable promoters such as the viral RNA-polymerase promoter are described in WO 95/16783 and WO 97/06250 and the clpP-promoter from Arabidopsis described in WO 99/46394.

[079] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to LMP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (1986, Antisense RNA as a molecular tool for genetic analysis, Reviews - Trends in Genetics, Vol. 1) and Mol et al. (1990, FEBS Lett. 268:427-430).

[080] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is to be understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a LMP can be expressed in bacterial cells, insect cells, fungal cells, mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells), algae, ciliates or plant cells. Other suitable host cells are known to those skilled in the art.

[081] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells including plant cells can be found in Sambrook et al. (1989, *Molecular Cloning: A Laboratory Manual. 2nd,*

ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) and other laboratory manuals such as Methods in Molecular Biology 1995, Vol. 44, Agrobacterium protocols, ed: Gartland and Davey, Humana Press, Totowa, New Jersey.

[082] For stable transfection of mammalian and plant cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin, kanamycin and methotrexate or in plants that confer resistance towards an herbicide such as glyphosate or glufosinate. A nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a LMP or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[083] To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of a LMP gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the LMP gene. Preferably, this LMP gene is an *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* LMP gene, but it can be a homologue from a related plant or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous LMP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a knock-out vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous LMP gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous LMP). To create a point mutation via homologous recombination, DNA-RNA hybrids can be used in a technique known as chimeraplasty (Cole-Strauss et al. 1999, Nucleic Acids Res. 27:1323-1330 and Kmiec 1999, American Scientist 87:240-247). Homologous recombination procedures in *Arabidopsis thaliana* are also well known in the art and are contemplated for use herein.

[084] In a homologous recombination vector, the altered portion of the LMP gene is flanked at its 5' and 3' ends by additional nucleic acid of the LMP gene to allow for homologous recombination to occur between the exogenous LMP gene carried by the vector and an endogenous LMP gene in a microorganism or plant. The additional flanking LMP nucleic

acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several hundreds of base pairs up to kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas & Capecchi 1987, Cell 51:503, for a description of homologous recombination vectors). The vector is introduced into a microorganism or plant cell (e.g., via polyethyleneglycol mediated DNA). Cells in which the introduced LMP gene has homologously recombined with the endogenous LMP gene are selected using art-known techniques.

[085] In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of a LMP gene on a vector placing it under control of the lac operon permits expression of the LMP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

[086] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture can be used to produce (i.e., express) a LMP. Accordingly, the invention further provides methods for producing LMPs using the host cells of the invention. In one embodiment, the method comprises culturing a host cell of the invention (into which a recombinant expression vector encoding a LMP has been introduced, or which contains a wild-type or altered LMP gene in it's genome) in a suitable medium until LMP is produced. In another embodiment, the method further comprises isolating LMPs from the medium or the host cell.

[087] Another aspect of the invention pertains to isolated LMPs, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of LMP in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of LMP having less than about 30% (by dry weight) of non-LMP (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-LMP, still more preferably less than about 10% of non-LMP, and most preferably less than about 5% non-LMP. When the LMP or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations

of LMP in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of LMP having less than about 30% (by dry weight) of chemical precursors or non-LMP chemicals, more preferably less than about 20% chemical precursors or non-LMP chemicals, still more preferably less than about 10% chemical precursors or non-LMP chemicals, and most preferably less than about 5% chemical precursors or non-LMP chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the LMP is derived. Typically, such proteins are produced by recombinant expression of, for example, an *Arabidopsis thaliana* and *Brassica napus* LMP in other plants than *Arabidopsis thaliana* and *Brassica napus* or microorganisms, algae or fungi.

[088] An isolated LMP or a portion thereof of the invention can participate in the metabolism of compounds necessary for the production of seed storage compounds in *Arabidopsis thaliana* and *Brassica napus*, or of cellular membranes, or has one or more of the activities set forth in Table 3. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence encoded by a nucleic acid of Appendix A such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *Arabidopsis thaliana* and *Brassica napus*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, a LMP of the invention has an amino acid sequence encoded by a nucleic acid of Appendix A. In yet another preferred embodiment, the LMP has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the LMP has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences encoded by a nucleic acid of Appendix A. The preferred LMPs of the present invention also preferably possess at least one of the LMP activities described herein. For example, a preferred LMP of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in

the metabolism of compounds necessary for the construction of cellular membranes in *Arabidopsis thaliana* and *Brassica napus*, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 3.

[089] In other embodiments, the LMP is substantially homologous to an amino acid sequence encoded by a nucleic acid of Appendix A and retains the functional activity of the protein of one of the sequences encoded by a nucleic acid of Appendix A yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail above. Accordingly, in another embodiment, the LMP is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence and which has at least one of the LMP activities described herein. In another embodiment, the invention pertains to a full *Arabidopsis thaliana* and *Brassica napus* protein which is substantially homologous to an entire amino acid sequence encoded by a nucleic acid of Appendix A.

[090] Dominant negative mutations or trans-dominant suppression can be used to reduce the activity of a LMP in transgenics seeds in order to change the levels of seed storage compounds. To achieve this a mutation that abolishes the activity of the LMP is created and the inactive non-functional LMP gene is overexpressed in the transgenic plant. The inactive trans-dominant LMP protein competes with the active endogenous LMP protein for substrate or interactions with other proteins and dilutes out the activity of the active LMP. In this way the biological activity of the LMP is reduced without actually modifying the expression of the endogenous LMP gene. This strategy was used by Pontier et al to modulate the activity of plant transcription factors (Pontier D, Miao ZH, Lam E, Plant J 2001 Sep;27(6):529-38, Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses).

[091] Homologues of the LMP can be generated by mutagenesis, e.g., discrete point mutation or truncation of the LMP. As used herein, the term "homologue" refers to a variant form of the LMP which acts as an agonist or antagonist of the activity of the LMP. An agonist of the LMP can retain substantially the same, or a subset, of the biological activities of the LMP. An antagonist of the LMP can inhibit one or more of the activities of the naturally occurring form of the LMP, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the LMP, or by binding to a LMP which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

[092] In an alternative embodiment, homologues of the LMP can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the LMP for LMP agonist or antagonist activity. In one embodiment, a variegated library of LMP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LMP variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LMP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of LMP sequences therein. There are a variety of methods which can be used to produce libraries of potential LMP homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential LMP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang 1983, Tetrahedron 39:3; Itakura et al. 1984, Annu. Rev. Biochem. 53:323; Itakura et al. 1984, Science 198:1056; Ike et al. 1983, Nucleic Acids Res. 11:477).

[093] In addition, libraries of fragments of the LMP coding sequences can be used to generate a variegated population of LMP fragments for screening and subsequent selection of homologues of a LMP. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a LMP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the LMP.

[094] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LMP homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into

replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LMP homologues (Arkin & Yourvan 1992, Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. 1993, Protein Engineering 6:327-331).

[095] In another embodiment, cell based assays can be exploited to analyze a variegated LMP library, using methods well known in the art.

[096] The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *Arabidopsis thaliana* and *Brassica napus* and related organisms; mapping of genomes of organisms related to *Arabidopsis thaliana* and *Brassica napus*; identification and localization of *Arabidopsis thaliana* and *Brassica napus* sequences of interest; evolutionary studies; determination of LMP regions required for function; modulation of a LMP activity; modulation of the metabolism of one or more cell functions; modulation of the transmembrane transport of one or more compounds; and modulation of seed storage compound accumulation.

[097] The plant *Arabidopsis thaliana* represents one member of higher (or seed) plants. It is related to other plants such as *Brassica napus* or soybean which require light to drive photosynthesis and growth. Plants like *Arabidopsis thaliana* and *Brassica napus* share a high degree of homology on the DNA sequence and polypeptide level, allowing the use of heterologous screening of DNA molecules with probes evolving from other plants or organisms, thus enabling the derivation of a consensus sequence suitable for heterologous screening or functional annotation and prediction of gene functions in third species. The ability to identify such functions can therefore have significant relevance, e.g., prediction of substrate specificity of enzymes. Further, these nucleic acid molecules may serve as reference points for the mapping of Arabidopsis genomes, or of genomes of related organisms.

[098] The LMP nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* or a close relative thereof. Also, they may be used to identify the presence of *Arabidopsis thaliana*, *Brassica napus*, and *Physcomitrella patens* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid

sequences of a number of *Arabidopsis thaliana* and *Brassica napus* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of an *Arabidopsis thaliana* and *Brassica napus* gene which is unique to this organism, one can ascertain whether this organism is present.

[099] Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *Arabidopsis thaliana* and *Brassica napus* proteins. For example, to identify the region of the genome to which a particular *Arabidopsis thaliana* and *Brassica napus* DNA-binding protein binds, the *Arabidopsis thaliana* and *Brassica napus* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *Arabidopsis thaliana* and *Brassica napus*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related plants.

[0100] The LMP nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

[0101] Manipulation of the LMP nucleic acid molecules of the invention may result in the production of LMPs having functional differences from the wild-type LMPs. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

[0102] There are a number of mechanisms by which the alteration of a LMP of the invention may directly affect the accumulation of seed storage compounds. In the case of plants expressing LMPs, increased transport can lead to altered accumulation of compounds and/or solute partitioning within the plant tissue and organs which ultimately could be used to affect the accumulation of one or more seed storage compounds during seed development. An example is provided by Mitsukawa et al. (1997, Proc. Natl. Acad. Sci. USA 94:7098-7102), where over expression of an Arabidopsis high-affinity phosphate transporter gene in tobacco cultured cells enhanced cell growth under phosphate-limited conditions. Phosphate availability also affects significantly the production of sugars and metabolic intermediates (Hurry et al. 2000, Plant J. 24:383-396) and the lipid composition in leaves and roots (Härtel et al. 2000, Proc. Natl. Acad. Sci. USA 97:10649-10654). Likewise, the activity of the plant ACCase has been demonstrated to be regulated by phosphorylation (Savage & Ohlrogge 1999, Plant J. 18:521-527) and alterations in the activity of the kinases and phosphatases (LMPs) that act on the ACCase could lead to increased or decreased levels of seed lipid accumulation. Moreover, the presence of lipid kinase activities in chloroplast envelope membranes suggests that signal transduction pathways and/or membrane protein regulation occur in envelopes (see, e.g., Müller et al. 2000, J. Biol. Chem. 275:19475-19481 and literature cited therein). The *ABI1* and *ABI2* genes encode two protein serine/threonine phosphatases 2C, which are regulators in abscisic acid signaling pathway, and thereby in early and late seed development (e.g. Merlot et al. 2001, Plant J. 25:295-303). For more examples see also the section 'background of the invention'.

[0103] The present invention also provides antibodies which specifically binds to an LMP-polypeptide, or a portion thereof, as encoded by a nucleic acid disclosed herein or as described herein.

[0104] Antibodies can be made by many well-known methods (see, e.g. Harlow and Lane, "Antibodies; A Laboratory Manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988). Briefly, purified antigen can be injected into an animal in an amount and in intervals sufficient to elicit an immune response. Antibodies can either be purified directly, or spleen cells can be obtained from the animal. The cells can then fused with an immortal cell line and screened for antibody secretion. The antibodies can be used to screen nucleic acid clone libraries for cells secreting the antigen. Those positive clones can then be sequenced (see, for example, Kelly et al. 1992, Bio/Technology 10:163-167; Bebbington et al. 1992, Bio/Technology 10:169-175).

[0105] The phrase "selectively binds" with the polypeptide refers to a binding reaction which is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bound to a particular protein do not bind in a significant amount to other proteins present in the sample. Selective binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies that selectively bind with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies selectively immunoreactive with a protein. See Harlow and Lane "Antibodies, A Laboratory Manual" Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that could be used to determine selective binding.

[0106] In some instances, it is desirable to prepare monoclonal antibodies from various hosts. A description of techniques for preparing such monoclonal antibodies may be found in Stites et al., editors, "Basic and Clinical Immunology," (Lange Medical Publications, Los Altos, Calif., Fourth Edition) and references cited therein, and in Harlow and Lane ("Antibodies, A Laboratory Manual" Cold Spring Harbor Publications, New York, 1988).

[0107] Throughout this application, various publications are referenced. The disclosures of all of these publications and those references cited within those publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

[0108] It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and Examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the claims included herein.

EXAMPLES

Example 1

General Processes

a) General Cloning Processes:

[0109] Cloning processes such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and

nylon membranes, linkage of DNA fragments, transformation of Escherichia coli and yeast cells, growth of bacteria and sequence analysis of recombinant DNA were carried out as described in Sambrook et al. (1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) or Kaiser, Michaelis and Mitchell (1994, "Methods in Yeast Genetics," Cold Spring Harbor Laboratory Press: ISBN 0-87969-451-3).

b) Chemicals:

[0110] The chemicals used were obtained, if not mentioned otherwise in the text, in p.a. quality from the companies Fluka (Neu-Ulm), Merck (Darmstadt), Roth (Karlsruhe), Serva (Heidelberg), and Sigma (Deisenhofen). Solutions were prepared using purified, pyrogen-free water, designated as H_2O in the following text, from a Milli-Q water system water purification plant (Millipore, Eschborn). Restriction endonucleases, DNA-modifying enzymes, and molecular biology kits were obtained from the companies AGS (Heidelberg), Amersham (Braunschweig), Biometra (Göttingen), Boehringer (Mannheim), Genomed (Bad Oeynnhausen), New England Biolabs (Schwalbach/ Taunus), Novagen (Madison, Wisconsin, USA), Perkin-Elmer (Weiterstadt), Pharmacia (Freiburg), Qiagen (Hilden), and Stratagene (Amsterdam, Netherlands). They were used, if not mentioned otherwise, according to the manufacturer's instructions.

c) Plant Material:

***Arabidopsis pkl* mutant**

[0111] For this study, in one series of experiments, root material of wild-type and pickle mutant *Arabidopsis thaliana* plants were used. The *pkl* mutation was isolated from an ethyl methanesulfonate-mutagenized population of the Columbia ecotype as described (Ogas et al., 1997, Science 277:91-94; Ogas et al., 1999, Proc. Natl. Acad. Sci. USA 96:13839-13844). In other series of experiments, siliques of individual ecotypes of *Arabidopsis thaliana* and of selected Arabidopsis phytohormone mutants were used. Seeds were obtained from the Arabidopsis stock center.

Brassica napus AC Excel and Cresor varieties

[0112] *Brassica napus* varieties AC Excel and Cresor were used for this study to create cDNA libraries. Seed, seed pod, flower, leaf, stem, and root tissues were collected from plants that were in some cases dark-, salt-, heat-,and drought-treated. However, this study focused on the use of seed and seed pod tissues for cDNA libraries.

d) Plant Growth:

Arabidopsis thaliana

[0113] Plants were either grown on Murashige-Skoog medium as described in Ogas et al. (1997, Science 277:91-94; 1999, Proc. Natl. Acad. Sci. USA 96:13839-13844) or on soil under standard conditions as described in Focks & Benning (1998, Plant Physiol. 118:91-101).

Brassica napus

[0114] Plants (AC Excel, except where mentioned) were grown in Metromix (Scotts, Marysville, OH) at 22°C under a 14/10 light/dark cycle. Six seed and seed pod tissues of interest in this study were collected to create the following cDNA libraries: Immature seeds, mature seeds, immature seed pods, mature seed pods, night-harvested seed pods, and Cresor variety (high erucic acid) seeds. Tissue samples were collected within specified time points for each developing tissue and multiple samples within a time frame pooled together for eventual extraction of total RNA. Samples from immature seeds were taken between 1-25 days after anthesis (daa), mature seeds between 25-50 daa, immature seed pods between 1-15 daa, mature seed pods between 15-50 daa, night-harvested seed pods between 1-50 daa and Cresor seeds 5-25 daa.

Example 2

Total DNA Isolation from Plants

[0115] The details for the isolation of total DNA relate to the working up of one gram fresh weight of plant material.

[0116] CTAB buffer: 2% (w/v) N-cethyl-N,N,N-trimethylammonium bromide (CTAB); 100 mM Tris HCl pH 8.0; 1.4 M NaCl; 20 mM EDTA. N-Laurylsarcosine buffer:10% (w/v) N-laurylsarcosine; 100 mM Tris HCl pH 8.0; 20 mM EDTA.

[0117] The plant material was triturated under liquid nitrogen in a mortar to give a fine powder and transferred to 2 ml Eppendorf vessels. The frozen plant material was then covered with a layer of 1 ml of decomposition buffer (1 ml CTAB buffer, 100 µl of N-laurylsarcosine buffer, 20 µl of β-mercaptoethanol and 10 µl of proteinase K solution, 10 mg/ml) and incubated at 60°C for one hour with continuous shaking. The homogenate obtained was distributed into two Eppendorf vessels (2 ml) and extracted twice by shaking with the same volume of chloroform/isoamyl alcohol (24:1). For phase separation, centrifugation was carried out at 8000*g* and RT for 15 minutes in each case. The DNA was then precipitated at -70°C for 30 minutes using ice-cold isopropanol. The precipitated DNA was sedimented at 4°C and 10,000 *g* for 30 minutes and resuspended in 180 µl of TE buffer (Sambrook et al., 1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6). For

further purification, the DNA was treated with NaCl (1.2 M final concentration) and precipitated again at -70°C for 30 minutes using twice the volume of absolute ethanol. After a washing step with 70% ethanol, the DNA was dried and subsequently taken up in 50 µl of H2O + RNAse (50 mg/ml final concentration). The DNA was dissolved overnight at 4°C and the RNAse digestion was subsequently carried out at 37°C for 1 hour. Storage of the DNA took place at 4°C.

Example 3

Isolation of Total RNA and poly-(A)+ RNA from Plants

Arabidopsis thaliana

[0118] For the investigation of transcripts, both total RNA and poly-(A)+ RNA were isolated. RNA was isolated from siliques of Arabidopsis plants according to the following procedure:

[0119] RNA preparation from *Arabidopsis* seeds - "hot" extraction:

Buffers, enzymes, and solutions:

- 2M KCl
- Proteinase K
- Phenol (for RNA)
- Chloroform:Isoamylalcohol
  (Phenol:choloroform 1:1; pH adjusted for RNA)
- 4 M LiCl, DEPC-treated
- DEPC-treated water
- 3M NaOAc, pH 5, DEPC-treated
- Isopropanol
- 70% ethanol (made up with DEPC-treated water)
- Resuspension buffer:0.5% SDS, 10 mM Tris pH 7.5, 1 mM EDTA made up with DEPC-treated water as this solution can not be DEPC-treated
- Extraction Buffer:
  0.2M Na Borate
  30 mM EDTA
  30 mM EGTA
  1% SDS (250µl of 10% SDS-solution for 2.5ml buffer)
  1% Deoxycholate (25mg for 2,5ml buffer)
  2% PVPP (insoluble - 50mg for 2.5ml buffer)
  2% PVP 40K (50mg for 2.5ml buffer)
  10 mM DTT
  100 mM β-Mercaptoethanol (fresh, handle under fume hood - use 35µl of 14.3M solution for 5ml buffer)

Extraction

[0120] Extraction buffer was heated up to 80°C. Tissues were ground in liquid nitrogen-cooled mortar, and the tissue powder was transferred to a 1.5ml tube. Tissues should be kept frozen until buffer is added; the sample should be transferred with a pre-cooled spatula; and the tube should be kept in liquid nitrogen at all times. Then 350µl preheated extraction buffer was added (For 100mg tissue, buffer volume can be as much as 500µl for bigger samples) to

tube; samples were vortexed; and the tube was heated to 80°C for approximately 1 minute and then kept on ice. The samples were vortexed and ground additionally with electric mortar.

Digestion

[0121] Proteinase K (0.15mg/100mg tissue) was added, and the mixture was vortexed and then kept at 37°C for one hour.

First Purification

[0122] For purification, 27µl 2M KCl was added to the samples. The samples were chilled on ice for 10 minutes and then centrifuged at 12.000 rpm for 10 minutes at room temperature. The supernatant was transferred to a fresh, RNAase-free tube, and one phenol extraction was conducted, followed by a choloroform:isoamylalcohol extraction. One volume isopropanol to was added to the supernatant, and the mixture was chilled on ice for 10 minutes. RNA was pelleted by centrifugation (7000 rpm for 10 minutes at room temperature). Pellets were dissolved in 1 ml 4M LiCl solution by vortexing the mixture 10 to 15 minutes. RNA was pelleted by a 5 minute centrifugation.

Second Purification

[0123] The pellet was resuspended in 500µl Resuspension buffer. Then 500 µl of phenol was added, and the mixture was vortexed. Then, 250µl chloroform:isoamylalcohol was added; the mixture was vortexed and then centrifuged for 5 minutes. The supernatant was transferred to a fresh tube. The choloform:isoamylalcohol extraction was repeated until the interface was clear. The supernatant was transferred to a fresh tube and 1/10 volume 3M NaOAc, pH 5 and 600µl isopropanol were added. The mixture was kept at -20 for 20 minutes or longer. The RNA was pelleted by 10 minutes of centrifugation, and then the pellet was washed once with 70% ethanol. All remaining alcohol was removed before dissolving the pellet in 15 to 20 µl DEPC-treated water. The quantity and quality of the RNA was determined by measuring the absorbance of a 1:200 dilution at 260nm and 280nm. (40µg RNA/ml = 1 OD_{260})

[0124] RNA from roots of wild-type Arabidopsis and the *pickle* mutant of Arabidopsis was isolated as described (Ogas et al., 1997, Science 277:91-94; Ogas et al., 1999, Proc. Natl. Acad. Sci. USA 96:13839-13844).

[0125] The mRNA was prepared from total RNA, using the Amersham Pharmacia Biotech mRNA purification kit, which utilizes oligo(dT)-cellulose columns.

[0126] Isolation of Poly-(A)+ RNA was isolated using Dyna BeadsR (Dynal, Oslo, Norway) following the instructions of the manufacturer's protocol. After determination of

the concentration of the RNA or of the poly(A)+ RNA, the RNA was precipitated by addition of 1/10 volume of 3 M sodium acetate pH 4.6 and 2 volumes of ethanol and stored at -70°C.

Brassica napus

[0127] Seeds were separated from pods to create homogeneous materials for seed and seed pod cDNA libraries. Tissues were ground into fine powder under liquid nitrogen using a mortar and pestle and transferred to a 50 ml tube. Tissue samples were stored at –80 °C until extractions could be performed. Total RNA was extracted from tissues using RNeasy Maxi kit (Qiagen) according to manufacturer's protocol, and mRNA was processed from total RNA using Oligotex mRNA Purification System kit (Qiagen), also according to manufacturer's protocol. The mRNA was sent to Hyseq Pharmaceuticals Incorporated (Sunnyville, CA) for further processing of mRNA from each tissue type into cDNA libraries and for use in their proprietary processes in which similar inserts in plasmids are clustered based on hybridization patterns.

Example 4

cDNA Library Construction

[0128] For cDNA library construction, first strand synthesis was achieved using Murine Leukemia Virus reverse transcriptase (Roche, Mannheim, Germany) and oligo-d(T)-primers, second strand synthesis by incubation with DNA polymerase I, Klenow enzyme and RNAseH digestion at 12°C (2 hours), 16°C (1 hour) and 22°C (1 hour). The reaction was stopped by incubation at 65°C (10 minutes) and subsequently transferred to ice. Double stranded DNA molecules were blunted by T4-DNA-polymerase (Roche, Mannheim) at 37°C (30 minutes). Nucleotides were removed by phenol/chloroform extraction and Sephadex G50 spin columns. EcoRI adapters (Pharmacia, Freiburg, Germany) were ligated to the cDNA ends by T4-DNA-ligase (Roche, 12°C, overnight) and phosphorylated by incubation with polynucleotide kinase (Roche, 37°C, 30 minutes). This mixture was subjected to separation on a low melting agarose gel. DNA molecules larger than 300 base pairs were eluted from the gel, phenol extracted, concentrated on Elutip-D-columns (Schleicher and Schuell, Dassel, Germany) and were ligated to vector arms and packed into lambda ZAPII phages or lambda ZAP-Express phages using the Gigapack Gold Kit (Stratagene, Amsterdam, Netherlands) using material and following the instructions of the manufacturer.

[0129] Brassica cDNA libraries were generated at Hyseq Pharmaceuticals Incorporated (Sunnyville, CA) No amplification steps were used in the library production to retain expression information. Hyseq's genomic approach involves grouping the genes into

clusters and then sequencing representative members from each cluster. The cDNA libraries were generated from oligo dT column purified mRNA. Colonies from transformation of the cDNA library into *E. coli* were randomly picked and the cDNA insert were amplified by PCR and spotted on nylon membranes. A set of ^{33}P radiolabeled oligonucleotides were hybridized to the clones, and the resulting hybridization pattern determined to which cluster a particular clone belonged. The cDNA clones and their DNA sequences were obtained for use in overexpression in transgenic plants and in other molecular biology processes described herein.

Example 5

Identification of LMP Genes of Interest
***Arabidopsis thaliana pkl* mutant**

[0130] The *pickle* Arabidopsis mutant was used to identify LMP-encoding genes. The pickle mutant accumulates seed storage compounds, such as seed storage lipids and seed storage proteins, in the root tips (Ogas et al., 1997, Science 277:91-94; Ogas et al., 1999, Proc. Natl. Acad. Sci. USA 96:13839-13844). The mRNA isolated from roots of wild-type and pickle plants was used to create a subtracted and normalized cDNA library (SSH library) containing cDNAs that are only present in the pickle roots, but not in the wild-type roots. Clones from the SSH library were spotted onto nylon membranes and hybridized with radio-labeled *pickle* or wild-type root mRNA to ascertain that the SSH clones were more abundant in pickle roots compared to wild-type roots. These SSH clones were randomly sequenced and the sequences were annotated (See Example 9). Based on the expression levels and on these initial functional annotations (See Table 3), clones from the SSH library were identified as potential LMP-encoding genes.

[0131] To identify additional potential gene targets from the Arabidopsis pickle mutant, the Megasort™ and MPSS technologies of Lynx Therapeutics Inc. were used. MegaSort is a micro-bead technology that allows both the simultaneous collection of millions of clones on as many micro-beads (See Brenner et al., 1999, Proc. Natl. Acad. Sci. USA 97:1665-1670). Genes are identified based on their differential expression in wild-type and pickle Arabidopsis mutant roots. RNA and mRNA are isolated from wild-type and mutant roots using standard procedures. The MegaSort technology enables the identification of over- and under-expressed clones in two mRNA samples without prior knowledge of the genes and is thus useful to discover differentially expressed genes that can encode LMP proteins. The MPSS technology enables the quantitation of the abundance of mRNA

transcripts in mRNA samples (Brenner et al., Nat. Biotechnol. 18:630-4) and was used to obtain expression profiles of wild-type and pickle root mRNAs.

[0132] Other LMP candidate genes were identified by randomly selecting various Arabidopsis phytohormone mutants (e.g. mutants obtained from EMS treatment) from the Arabidopsis stock center. These mutants and control wild-type plants were grown under standard conditions in growth chambers and screened for the accumulation of seed storage compounds. Mutants showing altered levels of seed storage compounds were considered as having a mutation in a LMP candidate gene and were investigated further.

Brassica napus

[0133] RNA expression profile data was obtained from the Hyseq clustering process. Clones showing 75% or greater expression in seed libraries compared to the other tissue libraries were selected as LMP candidate genes. The *Brassica napus* clones were selected for overexpression in Arabidopsis based on their expression profile.

Example 6

Cloning of full-length cDNAs and orthologs of identified LMP genes

Arabidopsis thaliana

[0134] Full-length sequences of the *Arabidopsis thaliana* partial cDNAs (ESTs) that were identified in the SSH library and from MegaSort and MPSS EST sequencing were isolated by RACE PCR using the SMART RACE cDNA amplification kit from Clontech allowing both 5' and 3' rapid amplification of cDNA ends (RACE). The isolation of cDNAs and the RACE PCR protocol used were based on the manufacturer's conditions. The RACE product fragments were extracted from agarose gels with a QIAquick Gel Extraction Kit (Qiagen) and ligated into the TOPO pCR 2.1 vector (Invitrogen) following manufacturer's instructions. Recombinant vectors were transformed into TOP10 cells (Invitrogen) using standard conditions (Sambrook et al., 1989). Transformed cells were grown overnight at 37°C on LB agar containing 50 μg/ml kanamycin and spread with 40 μl of a 40 mg/ml stock solution of X-gal in dimethylformamide for blue-white selection. Single white colonies were selected and used to inoculate 3 ml of liquid LB containing 50 μg/ml kanamycin and grown overnight at 37°C. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Subsequent analyses of clones and restriction mapping was performed according to standard molecular biology techniques (Sambrook et al., 1989).

[0135] Gene sequences can be used to identify homologous or heterologous genes (orthologs, the same LMP gene from another plant) from cDNA or genomic libraries. This can be done

by designing PCR primers to conserved sequences identified by multiple sequence alignments. Orthologs are often identified by designing degenerate primers to full-length or partial sequences of genes of interest. Homologous genes (e.g. full-length cDNA clones) can be isolated via nucleic acid hybridization using, for example, cDNA libraries: Depending on the abundance of the gene of interest, 100,000 up to 1,000,000 recombinant bacteriophages are plated and transferred to nylon membranes. After denaturation with alkali, DNA is immobilized on the membrane by e. g. UV cross linking. Hybridization is carried out at high stringency conditions. Aqueous solution hybridization and washing is performed at an ionic strength of 1 M NaCl and a temperature of 68°C. Hybridization probes are generated by, e.g., radioactive (^{32}P) nick transcription labeling (High Prime, Roche, Mannheim, Germany). Signals are detected by autoradiography.

[0136] Partially homologous or heterologous genes that are related but not identical can be identified in a procedure analogous to the above-described procedure using low stringency hybridization and washing conditions. For aqueous hybridization, the ionic strength is normally kept at 1 M NaCl while the temperature is progressively lowered from 68 to 42°C.

[0137] Isolation of gene sequences with homology (or sequence identity/similarity) only in a distinct domain (for example 10-20 amino acids) can be carried out by using synthetic radiolabeled oligonucleotide probes. Radiolabeled oligonucleotides are prepared by phosphorylation of the 5-prime end of two complementary oligonucleotides with T4 polynucleotide kinase. The complementary oligonucleotides are annealed and ligated to form concatemers. The double stranded concatemers are than radiolabeled by, for example, nick transcription. Hybridization is normally performed at low stringency conditions using high oligonucleotide concentrations.

Oligonucleotide hybridization solution:
6 x SSC
0.01 M sodium phosphate
1 mM EDTA (pH 8)
0.5 % SDS
100 µg/ml denaturated salmon sperm DNA
0.1 % nonfat dried milk

[0138] During hybridization, temperature is lowered stepwise to 5-10°C below the estimated oligonucleotide T_m or down to room temperature followed by washing steps and autoradiography. Washing is performed with low stringency such as three washing steps using 4 x SSC. Further details are described by Sambrook et al. (1989, "Molecular Cloning:

A Laboratory Manual", Cold Spring Harbor Laboratory Press) or Ausubel et al. (1994, "Current Protocols in Molecular Biology", John Wiley & Sons).

Brassica napus

[0139] Clones of *Brassica napus* genes obtained from Hyseq were sequenced at using a ABI 377 slab gel sequencer and BigDye Terminator Ready Reaction kits (PE Biosystems, Foster City, CA). Gene specific primers were designed using these sequences, and genes were amplified from the plasmid supplied from Hyseq using touch-down PCR. In some cases, primers were designed to add an "AACA" Kozak-like sequence just upstream of the gene start codon and two bases downstream were, in some cases, changed to GC to facilitate increased gene expression levels (Chandrashekhar et al., 1997, Plant Molecular Biology 35:993-1001). PCR reaction cycles were: 94°C, 5 minutes; 9 cycles of 94°C, 1 minute, 65°C, 1 minute, 72°C, 4 minutes and in which the anneal temperature was lowered by 1°C each cycle; 20 cycles of 94°C, 1 minute, 55°C, 1 minute, 72°C, 4 minutes; and the PCR cycle was ended with 72°C, 10 minutes. Amplified PCR products were gel purified from 1% agarose gels using GenElute –EtBr spin columns (Sigma), and after standard enzymatic digestion, were ligated into the plant binary vector pBPS-GB1 for transformation of Arabidopsis. The binary vector was amplified by overnight growth in *E. coli* DH5 in LB media and appropriate antibiotic, and plasmid was prepared for downstream steps using Qiagen MiniPrep DNA preparation kit. The insert was verified throughout the various cloning steps by determining its size through restriction digest and inserts were sequenced in parallel to plant transformations to ensure the expected gene was used in Arabidopsis transformation.

RT-PCR and cloning of Arabidopsis thaliana, Brassica napus, and Physcomitrella patens LMP genes

[0140] Full-length LMP cDNAs were isolated by RT-PCR from *Arabidopsis thaliana*, *Brassica napus*, or *Physcomitrella patens* RNA. The synthesis of the first strand cDNA was achieved using AMV Reverse Transcriptase (Roche, Mannheim, Germany). The resulting single-stranded cDNA was amplified via Polymerase Chain Reaction (PCR) utilizing two gene-specific primers. The conditions for the reaction were standard conditions with Expand High Fidelity PCR system (Roche). The parameters for the reaction were: five minutes at 94°C followed by five cycles of 40 seconds at 94°C, 40 seconds at 50°C, and 1.5 minutes at 72°C. This was followed by thirty cycles of 40 seconds at 94°C, 40 seconds at 65°C, and 1.5 minutes at 72°C. The fragments generated under these RT-PCR conditions were analyzed by

agarose gel electrophoresis to make sure that PCR products of the expected length had been obtained.

[0141] Full-length LMP cDNAs were isolated by using synthetic oligonucleotide primers (MWG-Biotech) designed based on the LMP gene specific DNA sequence that was determined by EST sequencing and by sequencing of RACE PCR products. The 5' PCR primers ("forward primer", F) for SEQ ID NO:83, SEQ ID NO:85, SEQ ID NO:87, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:97, SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, SEQ ID NO:105, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:111, SEQ ID NO:113, and SEQ ID NO:115 contained an AscI restriction site 5' upstream of the ATG start codon. The 5' PCR primers ("forward primer", F) for SEQ ID NO:117, SEQ ID NO:119, SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:151, SEQ ID NO:153, SEQ ID NO:155, SEQ ID NO:157, SEQ ID NO:159, SEQ ID NO:49, and SEQ ID NO:131, contained a NotI restriction site 5' upstream of the ATG start codon. The 3' PCR primers ("reverse primers", R) for SEQ ID NO:84, SEQ ID NO:86, SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:94, SEQ ID NO:96, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:106, SEQ ID NO:108, SEQ ID NO:110, SEQ ID NO:112, SEQ ID NO:114, and SEQ ID NO:116 contained a PacI restriction site 3' downstream of the stop codon. The 3' PCR primers ("reverse primers", R) for SEQ ID NO:118, SEQ ID NO:120, SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, and SEQ ID NO:140, contained a NotI restriction site 3' downstream of the stop codon. The 3' PCR primers ("reverse primers", R) for SEQ ID NO:142, SEQ ID NO:144, SEQ ID NO:146, SEQ ID NO:148, SEQ ID NO:150, SEQ ID NO:152, SEQ ID NO:156, SEQ ID NO:158, SEQ ID NO:160, SEQ ID NO:50, and SEQ ID NO:132, contained a StuI restriction site 3' downstream of the stop codon. The 3' PCR primers ("reverse primers", R) for SEQ ID NO:154 contained an EcoRV restriction site 3' downstream of the stop codon.

[0142] The restriction sites were added so that the RT-PCR amplification products could be cloned into the restriction sites located in the multiple cloning site of the binary vector. The following "forward" (F) and "reverse" (R) primers were used to amplify the full-length *Arabidopsis thaliana* or *Brassica napus* cDNAs by RT-PCR using RNA from *Arabidopsis thaliana* or *Brassica napus* as original template:

For amplification of SEQ ID NO:1

Pk123F (5'- ATGGCGCGCCATGGCAATCTTCCGAAGTACACTAGT-3') (SEQ ID NO:83)

Pk123R (5'- GCTTAATTAATTAAGGGCACTTGAGACGGCCA -3') (SEQ ID NO:84)

For amplification of SEQ ID NO:3

Pk197F (5'- ATGGCGCGCCAACAATGGAGAATGGAGCAACGACG -3') (SEQ ID NO:85)

Pk197R (5'- GCTTAATTAACTATATGGTTGGATATTGAGTCTTGGC -3') (SEQ ID NO:86)

For amplification of SEQ ID NO:5

Pk136F (5'- ATGGCGCGCCATGGCTGAAAAAGTAAAGTCTGGTCA-3') (SEQ ID NO:87)

Pk136R (5'- GCTTAATTAATTATAGCTCCTCAGATCCCTCCGA-3') (SEQ ID NO:88)

For amplification of SEQ ID NO:7

Pk156F (5'- ATGGCGCGCCATGGCTGGAGAAGAAATAGAGAGGG-3') (SEQ ID NO:89)

Pk156R (5'- GCTTAATTAATTAAACAGAGGCTTCTCTACTCTCACTT-3') (SEQ ID NO:90)

For amplification of SEQ ID NO:9

Pk159F (5'- ATGGCGCGCCATGGCTGGAGTGATGAAGTTGGC-3') (SEQ ID NO:91)

Pk159R (5'- GCTTAATTAATCACCTCACGGTGTTGCAGTTG-3') (SEQ ID NO:92)

For amplification of SEQ ID NO:11

Pk179F (5'-ATGGCGCGCCAAACAATGGGGCTTGCTGTGGTGG-3') (SEQ ID NO:93)

Pk179R (5'-GCTTAATTAATTACTGCAAGGCTTTCAATATATTTC-3') (SEQ ID NO:94)

For amplification of SEQ ID NO:13

Pk202F (5'- ATGGCGCGCCAACAATGGCGTTCACGGCGCTTGT-3') (SEQ ID NO:95)

Pk202R (5'- GCTTAATTAATCAACAAGTAGGATAAGGAACACCACA-3')
(SEQ ID NO:96)

For amplification of SEQ ID NO:15

Pk206F (5'- ATGGCGCGCCAACAATGGCCCTTGATGAGCTTCTCAAG-3')
(SEQ ID NO:97)
Pk206R (5'- GCTTAATTAATCAGAGAGAAGCAGAGTTTGTTCGC-3')
(SEQ ID NO:98)

For amplification of SEQ ID NO:17

Pk207F (5'- ATGGCGCGCCAACAATGGCGCAATCCCGATTATTAG-3')
(SEQ ID NO:99)
Pk207R (5'- GCTTAATTAATTAAAACCACTCGCCTCTCATTTC -3')
(SEQ ID NO:100)

For amplification of SEQ ID NO:19

Pk209F (5'- ATGGCGCGCCATGTCCGTGGCTCGATTCGAT -3')
(SEQ ID NO:101)
Pk209R (5'- GCTTAATTAACTAATCCTCTAGCTCGATGATTTTGAC-3')
(SEQ ID NO:102)

For amplification of SEQ ID NO:21

Pk215F (5'-ATGGCGCGCCAACAATGGCGATTTACAGATC
TCTAAGAAAG-3') (SEQ ID NO:103)
Pk215R (5'-GCTTAATTAATTACCTTAGATAAGTGATCCATGTCTGG-3')
(SEQ ID NO:104)

For amplification of SEQ ID NO:23

Pk239F (5'- ATGGCGCGCCAACAATGGTAAAGGAAACT
CTAATTCCTCCG-3') (SEQ ID NO:105)
Pk239R (5'-GCTTAATTAACTACCAGCCGAAGATTGGCTTGT-3')
(SEQ ID NO:106)

For amplification of SEQ ID NO:25

Pk240F (5'- ATGGCGCGCCATTTGGAGAGCAATGGCGACTT-3')
(SEQ ID NO:107)
Pk240R (5'- GCTTAATTAATTACATCGAACGAAGAAGC
ATCAA-3') (SEQ ID NO:108)

For amplification of SEQ ID NO:27

Pk241F (5'- ATGGCGCGCCCATCCTCAGAAAGAATGGCTCAAA-3')

(SEQ ID NO:109)

Pk241R (5'- GCTTAATTAATTAGCTTTCTTCACCATCATC

GGTG-3') (SEQ ID NO:110)

For amplification of SEQ ID NO:29

Pk242F (5'- ATGGCGCGCCAACAATGGGTGCAGGTGGAAGAATGCC-3')

(SEQ ID NO: 111)

Pk242R (5'- GCTTAATTAATCATAACTTATTGTTGTACCAGTA

CACACC-3') (SEQ ID NO:112)

For amplification of SEQ ID NO:31

Bn011F (5'- ATGGCGCGCCAACAATGGCTTCAATAAAT

GAAGATGTGTCT-3') (SEQ ID NO: 113)

Bn011R (5'- GACTTAATTAATCAATTGGTGGGATTAACGA

CTCCA-3') (SEQ ID NO:114)

For amplification of SEQ ID NO:33

Bn077F (5'-ATGGCGCGCCAACAATGGCTACA

TTCTCTTGTAATTCTTATGA-3') (SEQ ID NO: 115)

Bn077R (5'- GACTTAATTAATCAGAAGCGGCCATTAAAATT

ACCCA-3') (SEQ ID NO:116)

For amplification of SEQ ID NO:35

Jb001F (5'- ATAAGAATGCGGCCGCCATGGCAACGGAATGCATTGCA -3')

(SEQ ID NO:117)

Jb001R (5'- ATAAGAATGCGGCCGCTTAGAAACTTCT

TCTGTTCTT -3') (SEQ ID NO:118)

For amplification of SEQ ID NO:37

Jb002F (5'- ATAAGAATGCGGCCGCCATGGCGTCAGAGC

AAGCAAGG -3') (SEQ ID NO:119)

Jb002R (5'- ATAAGAATGCGGCCGCTCAACGTTGTCC

ATGTTCCCG -3') (SEQ ID NO:120)

For amplification of SEQ ID NO:39

Jb003F (5'- ATAAGAATGCGGCCGCCATGGCTAAGTC

TTGCTATTTCA -3') (SEQ ID NO:121)

Jb003R (5'- ATAAGAATGCGGCCGCTCAGGCGCTATAG

CCTAAGATT -3') (SEQ ID NO:122)

For amplification of SEQ ID NO:41

Jb005F (5'- ATAAGAATGCGGCCGCCATGGACGGTGCCGG
AGAATCACGA -3') (SEQ ID NO:123)
Jb005R (5'- ATAAGAATGCGGCCGCCTAATAACTTAA
AGTTACCGGA -3') (SEQ ID NO:124)

For amplification of SEQ ID NO:43

Jb007F (5'- ATAAGAATGCGGCCGCCATGTCGAGAGCTTTG
TCAGTCG -3') (SEQ ID NO:125)
Jb007R (5'- ATAAGAATGCGGCCGCCATGTCGAGAGCTTT
GTCAGTCG -3') (SEQ ID NO:126)

For amplification of SEQ ID NO:45

Jb009F (5'- ATAAGAATGCGGCCGCCATGGCAAGCAGCGAC
GTGAAGCT -3') (SEQ ID NO:127)
Jb009R (5'- ATAAGAATGCGGCCGCTCAACCAAGCCAAGAA
GCACCC -3') (SEQ ID NO:128)

For amplification of SEQ ID NO:47

Jb013F (5'- ATAAGAATGCGGCCGCCATGGCGTCTCAACAAGA
GAAGA -3') (SEQ ID NO:129)
Jb013R (5'- ATAAGAATGCGGCCGCTTAGGTCTTGGTCCTGA
ATTTG -3') (SEQ ID NO:130)

For amplification of SEQ ID NO:51

Jb017F (5'- ATAAGAATGCGGCCGCCATGGCTCCTTCAACAA
AAGTTC -3')(SEQ ID NO:133)
Jb017R (5'- ATAAGAATGCGGCCGCTCAAACACTGCTGATAGTATTT -3')
(SEQ ID NO:134)

For amplification of SEQ ID NO:53

Jb024F (5'- ATAAGAATGCGGCCGCCATGCGGTGCTTTCC
ACCTCCCT -3') (SEQ ID NO:135)
Jb024R (5'- ATAAGAATGCGGCCGCTTACTTTTGTAATGGTGAG
AGC -3') (SEQ ID NO:136)

For amplification of SEQ ID NO:55

Jb027F (5'- ATAAGAATGCGGCCGCCATGCTTCTAATTCTAG
CGATTT -3') (SEQ ID NO:137)
Jb027R (5'- ATAAGAATGCGGCCGCTCAGATAACCTTCTTCTTCTCG -3')
(SEQ ID NO:138)

For amplification of SEQ ID NO:57

OO-1F (5'- ATTGCGGCCGCACAATGGCACATGCCACGTTTACG -3')

(SEQ ID NO:139)

OO-1R (5'- ATTGCGGCCGCTTAGTCTTCATGGTCCCATAGATC -3')

(SEQ ID NO:140)

For amplification of SEQ ID NO:59

OO-2F (5'- GCGGCCGCCATGGCGTCTGAGAAACAAAAAC -3')

(SEQ ID NO:141)

OO-2R (5'- AGGCCTTTACGCATTTACCACAGCTCC -3') (SEQ ID NO:142)

For amplification of SEQ ID NO:61

OO-3F (5'- GCGGCCGCATGGATTCAACGAAGCTTAGTGAGC -3')

(SEQ ID NO:143)

OO-3R (5'- AGGCCTTTACTGAGGTCCTGCAAATTTG -3') (SEQ ID NO:144)

For amplification of SEQ ID NO:63

OO-4F (5'- GCGGCCGCCATGAAGGTTCACGAGACAAGA -3')

(SEQ ID NO:145)

OO-4R (5'- AGGCCTCTACTCTGGTTCGACATCGAC -3') (SEQ ID NO:146)

For amplification of SEQ ID NO:65

OO-5F (5'- GCGGCCGCCATGTCTACCCCAGCTGAATC -3') (SEQ ID NO:147)

OO-5R (5'- AGGCCTCTAATTGTAGAGATCATCATC -3') (SEQ ID NO:148)

For amplification of SEQ ID NO:67

OO-6F (5'- GCGGCCGCCATGGACAAATCTAGTACCATG -3')

(SEQ ID NO:149)

OO-6R (5'- AGGCCTTCAGCTACCACCCTTTTGTTTGAG -3') (SEQ ID NO:150)

For amplification of SEQ ID NO:69

OO-8F (5'- GCGGCCGCCATGGCGAAATCTCAGATCTGG -3')

(SEQ ID NO:151)

OO-8R (5'- AGGCCTTTAAGAAGAAGCAACGAACGTG -3') (SEQ ID NO:152)

For amplification of SEQ ID NO:71

OO-9F (5'- GCGGCCGCCATGGCGTCGAGCGATGAGCG -3') (SEQ ID NO:153)

OO-9R (5'- GATATCTTACGGGAACGGAGCCAATTTC -3') (SEQ ID NO:154)

For amplification of SEQ ID NO:73

OO-10F (5'- GCGGCCGCCATGGCGACTCTTAAGGTTTCTG -3')

(SEQ ID NO:155)

OO-10R (5'- AGGCCTTTAAGCATCATCTTCACCGAG -3') (SEQ ID NO:156)

For amplification of SEQ ID NO:75

OO-11F (5'- GCGGCCGCCATGGTGGATCTATTGAACTCG -3')

(SEQ ID NO:157)

OO-11R (5'- AGGCCTTTACAACTCTTGGATATTAAAC -3') (SEQ ID NO:158)

For amplification of SEQ ID NO:77

OO-12F (5'- GCGGCCGCCATGGCTGGAAAACTCATGCAC -3')

(SEQ ID NO:159)

OO-12R (5'- AGGCCTTTATGGCTCGACAATGATCTTC -3') (SEQ ID NO:160)

For amplification of SEQ ID NO:79

pp82F (5'- ATGGCGCGCCCGACATGAAGCGACGTTGAACG -3')

(SEQ ID NO:49)

pp82R (5'- GCTTAATTAACTTTCCGCAGCCTTCAGGCCGC -3')

(SEQ ID NO:50)

For amplification of SEQ ID NO:81

Pk225F (5'- GGTTAATTAAGGCGCGCCCCCGGAAGCGATGCTGAG -3')

(SEQ ID NO:131)

Pk225R (5'- ATCTCGAGGACGTCCCACAGCCACCGGATTC -3')

(SEQ ID NO:132)

Example 7

Identification of Genes of Interest by Screening Expression Libraries with Antibodies

[0143] The cDNA clones can be used to produce recombinant protein, for example, in *E. coli* (e. g. Qiagen QIAexpress pQE system). Recombinant proteins are then normally affinity purified via Ni-NTA affinity chromatography (Qiagen). Recombinant proteins can be used to produce specific antibodies for example by using standard techniques for rabbit immunization. Antibodies are affinity purified using a Ni-NTA column saturated with the recombinant antigen as described by Gu et al. (1994, BioTechniques 17:257-262). The antibody can then be used to screen expression cDNA libraries to identify homologous or heterologous genes via an immunological screening (Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press; or Ausubel et al. 1994, "Current Protocols in Molecular Biology", John Wiley & Sons).

Example 8

Northern-Hybridization

[0144] For RNA hybridization, 20 μg of total RNA or 1 μg of poly-(A)+ RNA was separated by gel electrophoresis in 1.25% strength agarose gels using formaldehyde as described in Amasino (1986, Anal. Biochem. 152:304), transferred by capillary attraction using 10 x SSC to positively charged nylon membranes (Hybond N+, Amersham, Braunschweig), immobilized by UV light, and pre-hybridized for 3 hours at 68°C using hybridization buffer (10% dextran sulfate w/v, 1 M NaCl, 1% SDS, 100 μg/ml of herring sperm DNA). The labeling of the DNA probe with the Highprime DNA labeling kit (Roche, Mannheim, Germany) was carried out during the pre-hybridization using alpha-^{32}P dCTP (Amersham, Braunschweig, Germany). Hybridization was carried out after addition of the labeled DNA probe in the same buffer at 68°C overnight. The washing steps were carried out twice for 15 minutes using 2 x SSC and twice for 30 minutes using 1 x SSC, 1% SDS at 68°C. The exposure of the sealed filters was carried out at -70°C for a period of 1 day to 14 days.

Example 9

DNA Sequencing and Computational Functional Analysis

[0145] The SSH cDNA library as described in Examples 4 and 5 was used for DNA sequencing according to standard methods, in particular by the chain termination method using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Weiterstadt, Germany). Random sequencing was carried out subsequent to preparative plasmid recovery from cDNA libraries via *in vivo* mass excision, retransformation, and subsequent plating of DH10B on agar plates (material and protocol details from Stratagene, Amsterdam, Netherlands). Plasmid DNA was prepared from overnight grown *E. coli* cultures grown in Luria-Broth medium containing ampicillin (See Sambrook et al. (1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6)) on a Qiagene DNA preparation robot (Qiagen, Hilden) according to the manufacturer's protocols. Sequencing primers with the following nucleotide sequences were used:

5′-CAGGAAACAGCTATGACC-3′	SEQ ID NO:161
5′-CTAAAGGGAACAAAAGCTG-3′	SEQ ID NO:162
5′-TGTAAAACGACGGCCAGT-3′	SEQ ID NO:163

[0146] Sequences were processed and annotated using the software package EST-MAX commercially provided by Bio-Max (Munich, Germany). The program incorporates

practically all bioinformatics methods important for functional and structural characterization of protein sequences. For reference see *http://pedant.mips.biochem.mpg.de*.

[0147] The most important algorithms incorporated in EST-MAX are: FASTA: Very sensitive protein sequence database searches with estimates of statistical significance (Pearson W.R., 1990, Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183:63-98); BLAST: Very sensitive protein sequence database searches with estimates of statistical significance (Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. Basic local alignment search tool. J. Mol. Biol. 215:403-410). PREDATOR: High-accuracy secondary structure prediction from single and multiple sequences. (Frishman & Argos 1997, 75% accuracy in protein secondary structure prediction. Proteins 27:329-335). CLUSTALW: Multiple sequence alignment (Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22:4673-4680). TMAP: Transmembrane region prediction from multiply aligned sequences (Persson B. & Argos P. 1994, Prediction of transmembrane segments in proteins utilizing multiple sequence alignments, J. Mol. Biol. 237:182-192). ALOM2:Transmembrane region prediction from single sequences (Klein P., Kanehisa M., and DeLisi C. 1984, Prediction of protein function from sequence properties: A discriminant analysis of a database. Biochim. Biophys. Acta 787:221-226. Version 2 by Dr. K. Nakai). PROSEARCH: Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M. and Smith J.E. 1992, ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. Biotechniques 13:919-921). BLIMPS: Similarity searches against a database of ungapped blocks (Wallace & Henikoff 1992, PATMAT:A searching and extraction program for sequence, pattern and block queries and databases, CABIOS 8:249-254. Written by Bill Alford).

Example 10

Plasmids for Plant Transformation

[0148] For plant transformation, various binary vectors such as a pBPS plant binary vector were used. Construction of the plant binary vectors was performed by ligation of the cDNA in sense or antisense orientation into the vector. In such vectors, a plant promoter was located 5-prime to the cDNA, where it activated transcription of the cDNA; and a polyadenylation sequence was located 3′-prime to the cDNA. Various plant promoters were used such as a constitutive promoter (Superpromoter), a seed-specific promoter, and a root-

specific promoter. Tissue-specific expression was achieved by using a tissue-specific promoter. For example, in some instances, seed-specific expression was achieved by cloning the napin or LeB4 or USP promoter 5-prime to the cDNA. Also, any other seed specific promoter element can be used, and such promoters are well known to one of ordinary skill in the art. For constitutive expression within the whole plant, in some instances, the Superpromoter or the CaMV 35S promoter was used. The expressed protein also can be targeted to a cellular compartment using a signal peptide, for example for plastids, mitochondria, or endoplasmic reticulum (Kermode, 1996, Crit. Rev. Plant Sci. 15:285-423). The signal peptide is cloned 5-prime in frame to the cDNA to achieve subcellular localization of the fusion protein.

[0149] The plant binary vectors comprised a selectable marker gene driven under the control of one of various plant promoters, such as the AtAct2-I promoter and the Nos-promoter; the LMP candidate cDNA under the control of a root-specific promoter, a seed-specific promoter, a non-tissue specific promoter, or a constitutive promoter; and a terminator. Partial or full-length LMP cDNA was cloned into the plant binary vector in sense or antisense orientation behind the desired promoter. The recombinant vector containing the gene of interest was transformed into Top10 cells (Invitrogen) using standard conditions. Transformed cells were selected for on LB agar containing the selective agent, and cells were grown overnight at 37°C. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Analysis of subsequent clones and restriction mapping was performed according to standard molecular biology techniques (Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY).

Example 11

Agrobacterium Mediated Plant Transformation

[0150] *Agrobacterium* mediated plant transformation with the LMP nucleic acids described herein can be performed using standard transformation and regeneration techniques (Gelvin, Stanton B. & Schilperoort R.A, Plant Molecular Biology Manual, 2nd ed. Kluwer Academic Publ., Dordrecht 1995 in Sect., Ringbuc Zentrale Signatur:BT11-P; Glick, Bernard R. and Thompson, John E. Methods in Plant Molecular Biology and Biotechnology, S. 360, CRC Press, Boca Raton 1993). For example, *Agrobacterium* mediated transformation can be performed using the GV3 (pMP90) (Koncz & Schell, 1986, Mol. Gen. Genet. 204:383-396) or LBA4404 (Clontech) *Agrobacterium tumefaciens* strain.

[0151] *Arabidopsis thaliana* can be grown and transformed according to standard conditions (Bechtold, 1993, Acad. Sci. Paris. 316:1194-1199; Bent et al., 1994, Science 265:1856-1860). Additionally, rapeseed can be transformed with the LMR nucleic acids of the present invention via cotyledon or hypocotyl transformation (Moloney et al., 1989, Plant Cell Report 8:238-242; De Block et al., 1989, Plant Physiol. 91:694-701). Use of antibiotics for *Agrobacterium* and plant selection depends on the binary vector and the *Agrobacterium* strain used for transformation. Rapeseed selection is normally performed using kanamycin as selectable plant marker. Additionally, *Agrobacterium* mediated gene transfer to flax can be performed using, for example, a technique described by Mlynarova et al. (1994, Plant Cell Report 13:282-285).

[0152] Transformation of soybean can be performed using for example a technique described in EP 0424 047, U.S. Patent No. 5,322,783 (Pioneer Hi-Bred International) or in EP 0397 687, U.S. Patent No. 5,376,543 or U.S. Patent No. 5,169,770 (University Toledo). Soybean seeds are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05% (v/v) Tween for 20 minutes with continuous shaking. Then the seeds are rinsed four times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 6 to 39 hours. The seed coats are peeled off, and cotyledons are detached from the embryo axis. The embryo axis is examined to make sure that the meristematic region is not damaged. The excised embryo axes are collected in a half-open sterile Petri dish and air-dried to a moisture content less than 20% (fresh weight) in a sealed Petri dish until further use.

[0153] The method of plant transformation is also applicable to *Brassica* and other crops. In particular, seeds of canola are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05 % (v/v) Tween for 20 minutes, at room temperature with continuous shaking. Then, the seeds are rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 18 hours. The seed coats are removed and the seeds are air dried overnight in a half-open sterile Petri dish. During this period, the seeds lose approximately 85% of their water content. The seeds are then stored at room temperature in a sealed Petri dish until further use.

[0154] *Agrobacterium tumefaciens* culture is prepared from a single colony in LB solid medium plus appropriate antibiotics (e.g. 100 mg/l streptomycin, 50 mg/l kanamycin) followed by growth of the single colony in liquid LB medium to an optical density at 600 nm

of 0.8. Then, the bacteria culture is pelleted at 7000 rpm for 7 minutes at room temperature, and resuspended in MS (Murashige & Skoog, 1962, Physiol. Plant. 15:473-497) medium supplemented with 100 mM acetosyringone. Bacteria cultures are incubated in this pre-induction medium for 2 hours at room temperature before use. The axis of soybean zygotic seed embryos at approximately 44% moisture content are imbibed for 2 h at room temperature with the pre-induced *Agrobacterium* suspension culture. (The imbibition of dry embryos with a culture of *Agrobacterium* is also applicable to maize embryo axes).

[0155] The embryos are removed from the imbibition culture and are transferred to Petri dishes containing solid MS medium supplemented with 2% sucrose and incubated for 2 days, in the dark at room temperature. Alternatively, the embryos are placed on top of moistened (liquid MS medium) sterile filter paper in a Petri dish and incubated under the same conditions described above. After this period, the embryos are transferred to either solid or liquid MS medium supplemented with 500 mg/l carbenicillin or 300 mg/l cefotaxime to kill the agrobacteria. The liquid medium is used to moisten the sterile filter paper. The embryos are incubated during 4 weeks at 25°C, under 440 µmol $m^{-2}s^{-1}$ and 12 hours photoperiod. Once the seedlings have produced roots, they are transferred to sterile metromix soil. The medium of the *in vitro* plants is washed off before transferring the plants to soil. The plants are kept under a plastic cover for 1 week to favor the acclimatization process. Then the plants are transferred to a growth room where they are incubated at 25°C, under 440 µmol $m^{-2}s^{-1}$ light intensity and 12 h photoperiod for about 80 days.

[0156] Samples of the primary transgenic plants (T_0) are analyzed by PCR to confirm the presence of T-DNA. These results are confirmed by Southern hybridization wherein DNA is electrophoresed on a 1% agarose gel and transferred to a positively charged nylon membrane (Roche Diagnostics). The PCR DIG Probe Synthesis Kit (Roche Diagnostics) is used to prepare a digoxigenin-labeled probe by PCR as recommended by the manufacturer.

Example 12

In vivo Mutagenesis

[0157] *In vivo* mutagenesis of microorganisms can be performed by incorporation and passage of the plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp W.D. 1996, DNA repair mechanisms, in: *Escherichia coli* and

Salmonella, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener and Callahan, 1994, Strategies 7:32-34. Transfer of mutated DNA molecules into plants is preferably done after selection and testing in microorganisms. Transgenic plants are generated according to various examples within the exemplification of this document.

Example 13

Assessment of the mRNA Expression and Activity of a Recombinant Gene Product in the Transformed Organism

[0158] The activity of a recombinant gene product in the transformed host organism can be measured on the transcriptional level or/and on the translational level. A useful method to ascertain the level of transcription of the gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. 1988, Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information at least partially demonstrates the degree of transcription of the transformed gene. Total cellular RNA can be prepared from plant cells, tissues or organs by several methods, all well-known in the art, such as that described in Bormann et al. (1992, Mol. Microbiol. 6:317-326).

[0159] To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (See, for example, Ausubel et al. 1988, Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

[0160] The activity of LMPs that bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such LMP on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar H. et al., 1995, EMBO J. 14:3895-3904 and

references cited therein). Reporter gene test systems are well known and established for applications in both prokaryotic and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

[0161] The determination of activity of lipid metabolism membrane-transport proteins can be performed according to techniques such as those described in Gennis R.B. (1989 Pores, Channels and Transporters, in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, pp. 85-137, 199-234 and 270-322).

Example 14

In vitro Analysis of the Function of Arabidopsis thaliana and Brassica napus Genes in Transgenic Plants

[0162] The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M. & Webb, E.C., 1979, Enzymes. Longmans: London; Fersht, 1985, Enzyme Structure and Mechanism. Freeman: New York; Walsh, 1979, Enzymatic Reaction Mechanisms. Freeman:San Francisco; Price, N.C., Stevens, L., 1982, Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., 1994, Enzymkinetik, 2nd ed. VCH:Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, Enzymes. VCH:Weinheim, p. 352-363.

Example 15

Analysis of the Impact of Recombinant LMPs on the Production of a Desired Seed Storage Compound: Fatty Acid Production

[0163] The total fatty acid content of Arabidopsis seeds was determined by saponification of seeds in 0.5 M KOH in methanol at 80°C for 2 hours followed by LC-MS analysis of the free fatty acids. Total fatty acid content of seeds of control and transgenic plants was measured with bulked seeds (usually 5 mg seed weight) of a single plant. Three different types of controls have been used: Col-2 (Columbia-2, the Arabidopsis ecotype in which SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21,

SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:79, or SEQ ID NO:81 has been transformed), Col-0 (Columbia-0, the Arabidopsis ecotype in which SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ:ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, or SEQ ID NO:77 has been transformed), C-24 (an Arabidopsis ecotype found to accumulate high amounts of total fatty acids in seeds), and the BPS empty (without an LMP gene of interest) binary vector construct. The controls indicated in the tables below have been grown side by side with the transgenic lines. Differences in the total values of the controls are explained either by differences in the growth conditions, which were found to be very sensitive to small variations in the plant cultivation, or by differences in the standards added to quantify the fatty acid content. Because of the seed bulking, all values obtained with T2 seeds, and in part also with T3 seeds, are the result of a mixture of homozygous (for the gene of interest) and heterozygous events, implying that these data underestimate the LMP gene effect.

[0164] Table 5. Determination of the T2 seed total fatty acid content of transgenic lines of pk123 (containing SEQ ID NO:1). Shown are the means (± standard deviation). (Average mean values are shown ± standard deviation, number of individual measurements per plant line: 12-20; Col-2 is the Arabidopsis ecotype the LMP gene has been transformed in, C-24 is a high-oil Arabidopsis ecotype used as another control).

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.318 ± 0.022
Col-2 wild-type control	0.300 ± 0.023
Pk123 transgenic seeds	0.319 ± 0.024

[0165] Table 6. Determination of the T2 seed total fatty acid content of transgenic lines of pk197 (containing SEQ ID NO:3). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.371 ± 0.010
Col-2 wild-type control	0.353 ± 0.017
Col-2 empty vector control	0.347 ± 0.024
Pk197 transgenic seeds	0.366 ± 0.014

[0166] **Table 7.** Determination of the T2 seed total fatty acid content of transgenic lines of pk136 (containing SEQ ID NO:5). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.351 ± 0.052
Col-2 wild-type control	0.344 ± 0.026
Col-2 empty vector control	0.346 ± 0.019
Pk136 transgenic seeds	0.374 ± 0.033

[0167] **Table 8.** Determination of the T2 seed total fatty acid content of transgenic lines of pk156 (containing SEQ ID NO:7). Shown are the means (± standard deviation) of 6 individual plants per line each.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.400 ± 0.001
Col-2 wild-type control	0.369 ± 0.043
Pk156 transgenic seeds	0.389 ± 0.007

[0168] **Table 9.** Determination of the T2 seed total fatty acid content of transgenic lines of pk159 (containing SEQ ID NO:9). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.413 ± 0.019
Col-2 wild-type control	0.381 ± 0.019
Pk159 transgenic seeds	0.409 ± 0.008

[0169] **Table 10.** Determination of the T2 seed total fatty acid content of transgenic lines of pk179 (containing SEQ ID NO:11). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.400 ± 0.033
Col-2 wild-type control	0.339 ± 0.033
Col-2 empty vector control	0.357 ± 0.021
Pk179 transgenic seeds	0.384 ± 0.020

[0170] **Table 11.** Determination of the T2 seed total fatty acid content of transgenic lines of pk202 (containing SEQ ID NO:13). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.413 ± 0.019
Col-2 wild-type control	0.381 ± 0.019
Col-2 empty vector control	0.407 ± 0.020
Pk202 transgenic seeds	0.426 ± 0.033

[0171] **Table 12.** Determination of the T2 seed total fatty acid content of transgenic lines of pk206 (containing SEQ ID NO:15). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.422 ± 0.013
Col-2 wild-type control	0.354 ± 0.026
Col-2 empty vector control	0.388 ± 0.023
Pk206 transgenic seeds	0.414 ± 0.031

[0172] **Table 13.** Determination of the T2 seed total fatty acid content of transgenic lines of pk207 (containing SEQ ID NO:17). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.371 ± 0.010
Col-2 wild-type control	0.353 ± 0.017
Col-2 empty vector control	0.347 ± 0.024
Pk207 transgenic seeds	0.370 ± 0.009

[0173] **Table 14.** Determination of the T2 seed total fatty acid content of transgenic lines of pk209 (containing SEQ ID NO:19). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.400 ± 0.001
Col-2 wild-type control	0.369 ± 0.043
Pk209 transgenic seeds	0.397 ± 0.007

[0174] **Table 15.** Determination of the T2 seed total fatty acid content of transgenic lines of pk215 (containing SEQ ID NO:21). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.373 ± 0.045
Col-2 wild-type control	0.344 ± 0.026
Col-2 empty vector control	0.346 ± 0.019
Pk215 transgenic seeds	0.401 ± 0.014

[0175] **Table 16.** Determination of the T3 seed total fatty acid content of transgenic lines of pk239 (containing SEQ ID NO:23). Shown are the means (± standard deviation) of 14-20 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.334 ± 0.030
Col-2 empty vector control	0.301 ± 0.027
Pk239-2 transgenic seeds	0.335 ± 0.028
Pk239-9 transgenic seeds	0.335 ± 0.018
Pk239-18 transgenic seeds	0.331 ± 0.026
Pk239-20 transgenic seeds	0.343 ± 0.022

[0176] **Table 17.** Determination of the T3 seed total fatty acid content of transgenic lines of pk240 (containing SEQ ID NO:25). Shown are the means (± standard deviation) of 10-20 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.393 ± 0.037
Col-2 empty vector control	0.342 ± 0.024
Pk240-3 transgenic seeds	0.373 ± 0.033
Pk240-6 transgenic seeds	0.388 ± 0.015
Pk240-10 transgenic seeds	0.393 ± 0.025

[0177] Table 18. Determination of the T2 seed total fatty acid content of transgenic lines of pk241 (containing SEQ ID NO:27). Shown are the means (± standard deviation) of 10 (controls) and 30 (pk241) individual plants per line, respectively.

Genotype	g total fatty acids/g seed weight
Col-2 wild-type control	0.312 ± 0.033
Col-2 empty vector control	0.305 ± 0.025
Pk241 transgenic seeds	0.336 ± 0.032

[0178] Table 19. Determination of the T2 seed total fatty acid content of transgenic lines of Pk242 (containing SEQ ID NO:29). Shown are the means (± standard deviation) of 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-2 wild-type control	0.344 ± 0.016
Col-2 empty vector control	0.333 ± 0.040
Pk242 transgenic seeds	0.364 ± 0.008

[0179] Table 20. Determination of the T2 seed total fatty acid content of transgenic lines of Bn011 (containing SEQ ID NO:31). Shown are the means (± standard deviation) of 14-20 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.334 ± 0.028
Col-2 wild-type control	0.286 ± 0.039
Col-2 empty vector control	0.291 ± 0.034
Bn011 transgenic seeds	0.308 ± 0.030

[0180] Table 21. Determination of the T2 seed total fatty acid content of transgenic lines of Bn077 (containing SEQ ID NO:33). Shown are the means (± standard deviation) of 8-17 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.366 ± 0.056
Col-2 wild-type control	0.290 ± 0.047
Col-2 empty vector control	0.292 ± 0.038
Bn077 transgenic seeds	0.314 ± 0.032

[0181] **Table 22.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb001 (containing SEQ ID NO:35). Shown are the means (± standard deviation) of 3 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.241 ± 0.012
Jb001 transgenic seeds	0.274 ± 0.003

[0182] **Table 23.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb002 (containing SEQ ID NO:37). Shown are the means (± standard deviation) of 3 individual control plants and 5 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.191 ± 0.044
Jb002 transgenic seeds	0.273 ± 0.020

[0183] **Table 24.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb003 (containing SEQ ID NO:39). Shown are the means (± standard deviation) of 3 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.267 ± 0.011
Jb003 transgenic seeds	0.297 ± 0.030

[0184] **Table 25.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb005 (containing SEQ ID NO:41). Shown are the means (± standard deviation) of 3 individual control plants and 7 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.229 ± 0.021
Jb005 transgenic seeds	0.264 ± 0.010

[0185] **Table 26.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb007 (containing SEQ ID NO:43). Shown are the means (± standard deviation) of 3 individual control plants and 5 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.296 ± 0.017
Jb007 transgenic seeds	0.320 ± 0.002

[0186] **Table 27.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb009 (containing SEQ ID NO:45). Shown are the means (± standard deviation) of 3 individual control plants and 3 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.227 ± 0.016
Jb009 transgenic seeds	0.238 ± 0.004

[0187] **Table 28.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb013 (containing SEQ ID NO:47). Shown are the means (± standard deviation) of 3 individual control plants and 4 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.243 ± 0.011
Jb013 transgenic seeds	0.262 ± 0.007

[0188] **Table 29.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb017 (containing SEQ ID NO:51). Shown are the means (± standard deviation) of 3 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.231 ± 0.020
Jb017 transgenic seeds	0.269 ± 0.022

[0189] **Table 30.** Determination of the T2 seed total fatty acid content of transgenic lines of Jb027 (containing SEQ ID NO:55). Shown are the means (± standard deviation) of 3 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.235 ± 0.052
Jb027 transgenic seeds	0.282 ± 0.014

[0190] Table 31. Determination of the T2 seed total fatty acid content of transgenic lines of OO-1 (containing SEQ ID NO:57). Shown are the means (± standard deviation) of 3 individual control plants and 7 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.250 ± 0.009
OO-1 transgenic seeds	0.274 ± 0.007

[0191] Table 32. Determination of the T2 seed total fatty acid content of transgenic lines of OO-4 (containing SEQ ID NO:63). Shown are the means (± standard deviation) of 2 individual control plants and 4 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.329 ± 0.041
OO-4 transgenic seeds	0.380 ± 0.015

[0192] Table 33. Determination of the T2 seed total fatty acid content of transgenic lines of OO-8 (containing SEQ ID NO:69). Shown are the means (± standard deviation) of 4 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.379 ± 0.009
OO-8 transgenic seeds	0.411 ± 0.008

[0193] Table 34. Determination of the T2 seed total fatty acid content of transgenic lines of OO-9 (containing SEQ ID NO:71). Shown are the means (± standard deviation) of 3 individual control plants and 4 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.315 ± 0.020
OO-9 transgenic seeds	0.333 ± 0.006

[0194] **Table 35.** Determination of the T2 seed total fatty acid content of transgenic lines of OO-11 (containing SEQ ID NO:75). Shown are the means (± standard deviation) of 3 individual control plants and 2 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.264 ± 0.003
OO-11 transgenic seeds	0.278 ± 0.003

[0195] **Table 36.** Determination of the T2 seed total fatty acid content of transgenic lines of OO-12 (containing SEQ ID NO:77). Shown are the means (± standard deviation) of 3 individual control plants and 9 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.290 ± 0.010
OO-12 transgenic seeds	0.316 ± 0.008

[0196] **Table 37.** Determination of the T4 seed total fatty acid content of transgenic lines of pp82 (containing SEQ ID NO:79). Shown are the means (± standard deviation) of 17-20 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.436 ± 0.050
Col-2 wild-type control	0.380 ± 0.020
Col-2 empty vector control	0.378 ± 0.030
pp82-15-16 transgenic seeds	0.432 ± 0.040
pp82-15-19 transgenic seeds	0.437 ± 0.040
pp82-16-10 transgenic seeds	0.430 ± 0.040
pp82-9-14 transgenic seeds	0.449 ± 0.040

[0197] **Table 38.** Determination of the T4 seed total fatty acid content of transgenic lines of pk225 (containing SEQ ID NO:81). This particular gene has been down-regulated. Shown are the means (± standard deviation) of 17-20 individual plants per line.

Genotype	g total fatty acids/g seed weight
C-24 wild-type control	0.344 ± 0.048
Col-2 empty vector control	0.327 ± 0.031
Pk225-11-19 transgenic seeds	0.350 ± 0.041
Pk225-19-8 transgenic seeds	0.351 ± 0.021
Pk225-7-6 transgenic seeds	0.354 ± 0.037
Pk225-9-10 transgenic seeds	0.363 ± 0.042

Table 39. Determination of the T2 seed total fatty acid content of transgenic lines of OO-3 (containing SEQ ID NO:61). Shown are the means (± standard deviation) of 4 individual control plants and 6 individual plants per line.

Genotype	g total fatty acids/g seed weight
Col-0 empty vector control	0.365 ± 0.006
OO-3 transgenic seeds	0.388 ± 0.006

Example 16

Analysis of the Impact of Recombinant Proteins on the Production of a Desired Seed Storage Compound

[0198] The effect of the genetic modification in plants on a desired seed storage compound (such as a sugar, lipid or fatty acid) can be assessed by growing the modified plant under suitable conditions and analyzing the seeds or any other plant organ for increased production of the desired product (i.e., a lipid or a fatty acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (See, for example, Ullman, 1985, Encyclopedia of Industrial Chemistry, vol. A2, pp. 89-90 and 443-613, VCH: Weinheim; Fallon, A. et al., 1987, Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al., 1993, Product recovery and purification, Biotechnology, vol. 3, Chapter III, pp. 469-714, VCH: Weinheim; Belter, P.A. et al., 1988, Bioseparations: downstream processing for biotechnology, John

Wiley & Sons; Kennedy J.F. & Cabral J.M.S., 1992, Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz J.A. & Henry J.D., 1988, Biochemical separations in: Ulmann's Encyclopedia of Industrial Chemistry, Separation and purification techniques in biotechnology, vol. B3, Chapter 11, pp. 1-27, VCH: Weinheim; and Dechow F.J. 1989).

[0199] Besides the above-mentioned methods, plant lipids are extracted from plant material as described by Cahoon et al. (1999, Proc. Natl. Acad. Sci. USA 96, 22:12935-12940) and Browse et al. (1986, Anal. Biochemistry 442:141-145). Qualitative and quantitative lipid or fatty acid analysis is described in Christie, William W., Advances in Lipid Methodology. Ayr/Scotland :Oily Press. - (Oily Press Lipid Library; Christie, William W., Gas Chromatography and Lipids. A Practical Guide - Ayr, Scotland :Oily Press, 1989 Repr. 1992. - IX,307 S. - (Oily Press Lipid Library; and "Progress in Lipid Research, Oxford :Pergamon Press, 1 (1952) – 16 (1977) Progress in the Chemistry of Fats and Other Lipids CODEN.

[0200] Unequivocal proof of the presence of fatty acid products can be obtained by the analysis of transgenic plants following standard analytical procedures: GC, GC-MS or TLC as variously described by Christie and references therein (1997 in: Advances on Lipid Methodology 4th ed.: Christie, Oily Press, Dundee, pp. 119-169; 1998). Detailed methods are described for leaves by Lemieux et al. (1990, Theor. Appl. Genet. 80:234-240) and for seeds by Focks & Benning (1998, Plant Physiol. 118:91-101).

[0201] Positional analysis of the fatty acid composition at the C-1, C-2 or C-3 positions of the glycerol backbone is determined by lipase digestion (See, e.g., Siebertz & Heinz 1977, Z. Naturforsch. 32c:193-205, and Christie, 1987, Lipid Analysis 2nd Edition, Pergamon Press, Exeter, ISBN 0-08-023791-6).

[0202] A typical way to gather information regarding the influence of increased or decreased protein activities on lipid and sugar biosynthetic pathways is for example via analyzing the carbon fluxes by labeling studies with leaves or seeds using ^{14}C-acetate or ^{14}C-pyruvate (See, e.g. Focks & Benning, 1998, Plant Physiol. 118:91-101; Eccleston & Ohlrogge, 1998, Plant Cell 10:613-621). The distribution of carbon-14 into lipids and aqueous soluble components can be determined by liquid scintillation counting after the respective separation (for example on TLC plates) including standards like ^{14}C-sucrose and ^{14}C-malate (Eccleston & Ohlrogge, 1998, Plant Cell 10:613-621).

[0203] Material to be analyzed can be disintegrated via sonification, glass milling, liquid nitrogen and grinding, or via other applicable methods. The material has to be

centrifuged after disintegration. The sediment is resuspended in distilled water, heated for 10 minutes at 100°C, cooled on ice and centrifuged again, followed by extraction in 0.5 M sulfuric acid in methanol containing 2% dimethoxypropane for 1 hour at 90°C, leading to hydrolyzed oil and lipid compounds resulting in transmethylated lipids. These fatty acid methyl esters are extracted in petrolether and finally subjected to GC analysis using a capillary column (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 m, 0.32 mm) at a temperature gradient between 170°C and 240°C for 20 minutes and 5 minutes at 240°C. The identity of resulting fatty acid methylesters is defined by the use of standards available form commercial sources (e.g., Sigma).

[0204] In the case of fatty acids where standards are not available, molecule identity is shown via derivatization and subsequent GC-MS analysis. For example, the localization of triple bond fatty acids is shown via GC-MS after derivatization via 4,4-Dimethoxy-oxazolin-Derivaten (Christie, Oily Press, Dundee, 1998).

[0205] A common standard method for analyzing sugars, especially starch, is published by Stitt M., Lilley R.Mc.C., Gerhardt R. and Heldt M.W. (1989, "Determination of metabolite levels in specific cells and subcellular compartments of plant leaves," Methods Enzymol. 174:518-552; for other methods, see also Härtel et al., 1998, Plant Physiol. Biochem. 36:407-417 and Focks & Benning, 1998, Plant Physiol. 118:91-101).

[0206] For the extraction of soluble sugars and starch, 50 seeds are homogenized in 500 µl of 80% (v/v) ethanol in a 1.5-ml polypropylene test tube and incubated at 70°C for 90 minutes. Following centrifugation at 16,000 *g* for 5 minutes, the supernatant is transferred to a new test tube. The pellet is extracted twice with 500 µl of 80% ethanol. The solvent of the combined supernatants is evaporated at room temperature under a vacuum. The residue is dissolved in 50 µl of water, representing the soluble carbohydrate fraction. The pellet left from the ethanol extraction, which contains the insoluble carbohydrates including starch, is homogenized in 200 µl of 0.2 N KOH, and the suspension is incubated at 95°C for 1 hour to dissolve the starch. Following the addition of 35 µl of 1 N acetic acid and centrifugation for 5 minutes at 16,000 *g*, the supernatant is used for starch quantification.

[0207] To quantify soluble sugars, 10 µl of the sugar extract is added to 990 µl of reaction buffer containing 100 mM imidazole, pH 6.9, 5 mM $MgCl_2$, 2 mM NADP, 1 mM ATP, and 2 units 2 ml^{-1} of Glucose-6-P-dehydrogenase. For enzymatic determination of glucose, fructose, and sucrose, 4.5 units of hexokinase, 1 unit of phosphoglucoisomerase, and 2 µl of a saturated fructosidase solution are added in succession. The production of NADPH

is photometrically monitored at a wavelength of 340 nm. Similarly, starch is assayed in 30 µl of the insoluble carbohydrate fraction with a kit from Boehringer Mannheim.

[0208] An example for analyzing the protein content in leaves and seeds can be found by Bradford M.M. (1976, "A rapid and sensitive method for the quantification of microgram quantities of protein using the principle of protein dye binding," Anal. Biochem. 72:248-254). For quantification of total seed protein, 15-20 seeds are homogenized in 250 µl of acetone in a 1.5-ml polypropylene test tube. Following centrifugation at 16,000 *g*, the supernatant is discarded and the vacuum-dried pellet is resuspended in 250 µl of extraction buffer containing 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, and 1% (w/v) SDS. Following incubation for 2 h at 25°C, the homogenate is centrifuged at 16,000 *g* for 5 min and 200 ml of the supernatant will be used for protein measurements. In the assay, γ-globulin is used for calibration. For protein measurements, Lowry DC protein assay (Bio-Rad) or Bradford-assay (Bio-Rad) are used.

[0209] Enzymatic assays of hexokinase and fructokinase are performed spectrophotometrically according to Renz et al. (1993, Planta 190:156-165); enzymatic assays of phosphogluco-isomerase, ATP-dependent 6-phosphofructokinase, pyrophosphate-dependent 6-phospho-fructokinase, Fructose-1,6-bisphosphate aldolase, triose phosphate isomerase, glyceral-3-P dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase are performed according to Burrell et al. (1994, Planta 194:95-101); and enzymatic assays of UDP-Glucose-pyrophosphorylase according to Zrenner et al. (1995, Plant J. 7:97-107).

[0210] Intermediates of the carbohydrate metabolism, like Glucose-1-phosphate, Glucose-6-phosphate, Fructose-6-phosphate, Phosphoenolpyruvate, Pyruvate, and ATP are measured as described in Härtel et al. (1998, Plant Physiol. Biochem. 36:407-417), and metabolites are measured as described in Jelitto et al. (1992, Planta 188:238-244).

[0211] In addition to the measurement of the final seed storage compound (i.e., lipid, starch or storage protein), it is also possible to analyze other components of the metabolic pathways utilized for the production of a desired seed storage compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound (Fiehn et al., 2000, Nature Biotech. 18:1447-1161).

[0212] For example, yeast expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into *Saccharomyces*

cerevisiae using standard protocols. The resulting transgenic cells can then be assayed for alterations in sugar, oil, lipid, or fatty acid contents.

[0213] Similarly, plant expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into an appropriate plant cell such as Arabidopsis, soybean, rape, maize, wheat, *Medicago truncatula*, etc., using standard protocols. The resulting transgenic cells and/or plants derived therefrom can then be assayed for alterations in sugar, oil, lipid, or fatty acid contents.

[0214] Additionally, the sequences disclosed herein, or fragments thereof, can be used to generate knockout mutations in the genomes of various organisms, such as bacteria, mammalian cells, yeast cells, and plant cells (Girke at al., 1998, Plant J. 15:39-48). The resultant knockout cells can then be evaluated for their composition and content in seed storage compounds, and the effect on the phenotype and/or genotype of the mutation. For other methods of gene inactivation include US 6,004,804 and Puttaraju et al., 1999, Nature Biotech. 17:246-252).

Example 17

Purification of the Desired Product from Transformed Organisms

[0215] An LMP can be recovered from plant material by various methods well known in the art. Organs of plants can be separated mechanically from other tissue or organs prior to isolation of the seed storage compound from the plant organ. Following homogenization of the tissue, cellular debris is removed by centrifugation and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from cells grown in culture, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

[0216] The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

[0217] There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey J.E. & Ollis D.F., 1986, Biochemical Engineering Fundamentals, McGraw-Hill:New York.

[0218] The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, analytical chromatography such as high performance liquid chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994, Appl. Environ. Microbiol. 60:133-140), Malakhova et al. (1996, Biotekhnologiya 11:27-32), Schmidt et al. (1998, Bioprocess Engineer 19:67-70), Ulmann's Encyclopedia of Industrial Chemistry (1996, Vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587) and Michal G. (1999, Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. 1987, Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17).

Example 18

Screening for increased stress tolerance and plant growth

[0219] The transgenic plants are screened for their improved stress tolerance demonstrating that transgene expression confers stress tolerance. The transgenic plants are further screened for their growth rate demonstrating that transgene expression confers increased growth rates and/or increased seed yield.

[0220] Classification of the proteins was done by Blasting against the BLOCKS database (S. Henikoff & J. G. Henikoff, "Protein family classification based on searching a database of blocks", Genomics 19:97-107 (1994)).

[0221] Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompasses by the claims to the invention disclosed and claimed herein.

Appendix A

SEQ ID NO:1, Nucleotide sequence of the open reading frame of Pk123

ATGGCAATCTTCCGAAGTACACTAGTTTTACTGCTGATCCTCTTCTGCCTCACCAC
TTTTGAGCTTCATGTTCATGCTGCTGAAGATTCACAAGTCGGTGAAGGCGTAGTG
AAAATTGATTGCGGTGGGAGATGCAAAGGTAGATGCAGCAAATCGTCGAGGCCA
AATCTGTGTTTGAGAGCATGCAACAGCTGTTGTTACCGCTGCAACTGTGTGCCAC
CAGGCACCGCCGGGAACCACCACCTTTGTCCTTGCTACGCCTCCATTACCACTCG
TGGTGGCCGTCTCAAGTGCCCTTAA

SEQ ID NO:2, Deduced amino acid sequence of the open reading frame of Pk123

MAIFRSTLVLLLILFCLTTFELHVHAAEDSQVGEGVVKIDCGGRCKGRCSKSSRPNLC
LRACNSCCYRCNCVPPGTAGNHHLCPCYASITTRGGRLKCP

SEQ ID NO:3, Nucleotide sequence of the open reading frame of Pk197

ATGGAGAATGGAGCAACGACGACGAGCACAATTACCATCAAAGGGATTCTGAGT
TTGCTAATGGAAAGCATCACAACAGAGGAAGATGAAGGAGGAAAGAGAGTAAT
ATCTCTGGGAATGGGAGACCCAACACTCTACTCGTGTTTTCGTACAACACAAGTC
TCTCTTCAAGCTGTTTCTGATTCTCTTCTCTCCAACAAGTTCCATGGTTACTCTCCT
ACCGTCGGTCTTCCCCAAGCTCGAAGGGCAATAGCAGAGTATCTATCGCGTGATC
TTCCATACAAACTTTCACAGGATGATGTGTTTATCACATCGGGTTGCACGCAAGC
GATCGATGTAGCATTGTCGATGTTAGCTCGTCCCAGGGCTAATATACTTCTTCCA
AGGCCTGGTTTCCCAATCTATGAACTCTGTGCTAAGTTTAGACACCTTGAAGTTC
GCTACGTCGATCTTCTTCCGGAAAATGGATGGGAGATCGATCTTGATGCTGTCGA
GGCTCTTGCAGACGAAAACACGGTTGCTTTGGTTGTTATAAACCCTGGTAATCCT
TGCGGGAATGTCTATAGCTACCAGCATTTGATGAAGATTGCGGAATCGGCGAAA
AAACTAGGGTTTCTTGTGATTGCTGATGAGGTTTACGGTCATCTTGCTTTTGGTAG
CAAACCGTTTGTGCCAATGGGTGTGTTTGGATCTATTGTTCCTGTGCTTACTCTTG
GCTCTTTATCAAAGAGATGGATAGTTCCAGGTTGGCGACTCGGGTGGTTTGTCAC
CACTGATCCTTCTGGTTCCTTTAAGGACCCTAAGATCATTGAGAGGTTTAAGAAA
TACTTTGATATTCTTGGTGGACCAGCTACATTTATTCAGGCTGCAGTTCCCACTAT
TTTGGAACAGACGGATGAGTCTTTCTTCAAGAAAACCTTGAACTCGTTGAAGAAC
TCTTCGGATATTTGTTGTGACTGGATCAAGGAGATTCCTTGCATTGATTCCTCGCA
TCGACCAGAAGGATCCATGGCAATGATGGTTAAGCTGAATCTCTCATTACTTGAA
GATGTAAGTGACGATATCGACTTCTGTTTCAAGTTAGCTAGGGAAGAATCAGTCA
TCCTTCTTCCTGGGACCGCGGTGGGGCTGAAGAACTGGCTGAGGATAACGTTTGC
AGCAGATGCAACTTCGATTGAAGAAGCTTTTAAAAGGATCAAATGTTTCTATCTT
AGACATGCCAAGACTCAATATCCAACCATATAG

SEQ ID NO:4, Deduced amino acid sequence of the open reading frame of Pk197

MENGATTTSTITIKGILSLLMESITTEEDEGGKRVISLGMGDPTLYSCFRTTQVSLQAV
SDSLLSNKFHGYSPTVGLPQARRAIAEYLSRDLPYKLSQDDVFITSGCTQAIDVALSM
LARPRANILLPRPGFPIYELCAKFRHLEVRYVDLLPENGWEIDLDAVEALADENTVAL
VVINPGNPCGNVYSYQHLMKIAESAKKLGFLVIADEVYGHLAFGSKPFVPMGVFGSI
VPVLTLGSLSKRWIVPGWRLGWFVTTDPSGSFKDPKIIERFKKYFDILGGPATFIQAA
VPTILEQTDESFFKKTLNSLKNSSDICCDWIKEIPCIDSSHRPEGSMAMMVKLNLSLLE
DVSDDIDFCFKLAREESVILLPGTAVGLKNWLRITFAADATSIEEAFKRIKCFYLRHAK
TQYPTI

SEQ ID NO:5, Nucleotide sequence of the open reading frame of Pk136

ATGGCTGAAAAAGTAAAGTCTGGTCAAGTTTTTAACCTATTATGCATATTCTCGA
TCTTTTTCTTCCTCTTTGTGTTATCAGTGAATGTTTCGGCTGATGTCGATTCTGAGA
GAGCGGTGCCATCTGAAGATAAAACGACGACTGTTTGGCTAACTAAAATCAAAC
GGTCCGGTAAAAATTATTGGGCTAAAGTTAGAGAGACTTTGGATCGTGGACAGT
CCCACTTCTTTCCTCCGAACACATATTTTACCGGAAAGAATGATGCGCCGATGGG
AGCCGGTGAAAATATGAAAGAGGCGGCGACGAGGAGCTTTGAGCATAGCAAAG
CGACGGTGGAGGAAGCTGCTAGATCAGCGGCAGAAGTGGTGAGTGATACGGCGG
AAGCTGTGAAAGAAAAGGTGAAGAGGAGCGTTTCCGGTGGAGTGACGCAGCCGT
CGGAGGGATCTGAGGAGCTATAA

SEQ ID NO:6, Deduced amino acid sequence of the open reading frame of Pk136

MAEKVKSGQVFNLLCIFSIFFFLFVLSVNVSADVDSERAVPSEDKTTTVWLTKIKRSG
KNYWAKVRETLDRGQSHFFPPNTYFTGKNDAPMGAGENMKEAATRSFEHSKATVE
EAARSAAEVVSDTAEAVKEKVKRSVSGGVTQPSEGSEEL

SEQ ID NO:7, Nucleotide sequence of the open reading frame of Pk156

ATGGCTGGAGAAGAAATAGAGAGGGAGAAGAAATCTGCAGCATCTGCAAGAAC
TCACACCAGAAACAACACTCAACAAAGTTCTTCTTCTGGTTATCTGAAAACGCTT
CTCCTGGTAACGTTCGTCGGAGTTTTAGCATGGGTTTATCAAACAATCCAACCAC
CACCCGCCAAAATCGTCGGCTCTCCCGGTGGACCCACCGTGACATCACCGAGGAT
CAAACTGAGAGACGGAAGACATCTGGCTTACACAGAATTCGGAATCCCTAGAGA
CGAAGCCAAGTTCAAGATCATAAACATCCACGGCTTCGATTCTTGTATGCGAGAC
TCGCATTTCGCCAATTTCTTATCGCCGGCTCTTGTGGAGGAATTGAGGATATACA
TTGTGTCTTTTGATCGTCCTGGTTATGGAGAGAGTGATCCTAACCTGAATGGGTC
ACCAAGAAGCATAGCATTGGATATAGAAGAGCTTGCTGATGGGTTAGGACTAGG
ACCTCAGTTCTATCTCTTTGGTTACTCCATGGGTGGTGAAATTACATGGGCATGCC
TTAACTACATTCCTCACAGGTTAGCAGGAGCTGCCCTTGTAGCTCCAGCGATTAA
CTATTGGTGGAGAAACTTACCGGGAGATTTAACAAGAGAAGCTTTCTCTCTTATG
CATCCTGCAGATCAATGGTCACTTCGAGTAGCTCATTATGCTCCTTGGCTTACATA
TTGGTGGAACACTCAGAAATGGTTCCCAATCTCCAATGTGATTGCCGGTAATCCC
ATTATTTTCTCACGTCAGGACATGGAGATCTTGTCGAAGCTCGGATTCGTCAATC
CAAATCGGGCATACATAAGACAACAAGGTGAATATGTAAGCTTACACCGAGATT
TGAATGTCGCATTTTCAAGCTGGGAGTTTGATCCGTTAGACCTTCAAGATCCGTT
CCCGAACAACAATGGCTCAGTTCACGTATGGAATGGCGATGAGGATAAGTTTGT
GCCAGTAAAGCTTCAACGGTATGTCGCGTCAAAGCTGCCATGGATTCGTTACCAT
GAAATATCTGGATCAGGACATTTTGTACCATTTGTGGAAGGTATGACTGATAAGA
TCATCAAGTCACTTTTGGTTGGGGAAGAAGATGTAAGTGAGAGTAGAGAAGCCT
CTGTTTAA

SEQ ID NO:8, Deduced amino acid sequence of the open reading frame of Pk156

MAGEEIEREKKSAASARTHTRNNTQQSSSSGYLKTLLLVTFVGVLAWVYQTIQPPPA
KIVGSPGGPTVTSPRIKLRDGRHLAYTEFGIPRDEAKFKIINIHGFDSCMRDSHFANFLS
PALVEELRIYIVSFDRPGYGESDPNLNGSPRSIALDIEELADGLGLGPQFYLFGYSMGG
EITWACLNYIPHRLAGAALVAPAINYWWRNLPGDLTREAFSLMHPADQWSLRVAH
YAPWLTYWWNTQKWFPISNVIAGNPIIFSRQDMEILSKLGFVNPNRAYIRQQGEYVS
LHRDLNVAFSSWEFDPLDLQDPFPNNNGSVHVWNGDEDKFVPVKLQRYVASKLPWI
RYHEISGSGHFVPFVEGMTDKIIKSLLVGEEDVSESREASV

SEQ ID NO:9, Nucleotide sequence of the open reading frame of Pk159

ATGGCTGGAGTGATGAAGTTGGCATGCATGGTCTTGGCTTGCATGATTGTGGCCG
GTCCAATCACAGCGAACGCGCTTATGAGTTGTGGCACCGTCAACGGCAACCTGG
CAGGGTGCATTGCCTACTTGACCCGAGGTGCTCCACTTACCCAAGGGTGCTGCAA
CGGCGTTACTAACCTTAAAAACATGGCCAGTACAACCCCAGACCGTCAGCAAGC
TTGCCGTTGCCTTCAATCTGCCGCTAAAGCCGTTGGTCCCGGTCTCAACACTGCCC
GTGCAGCTGGACTTCCTAGCGCATGCAAAGTCAATATTCCTTACAAAATCAGCGC
CAGCACCAACTGCAACACCGTGAGGTGA

SEQ ID NO:10, Deduced amino acid sequence of the open reading frame of Pk159

MAGVMKLACMVLACMIVAGPITANALMSCGTVNGNLAGCIAYLTRGAPLTQGCCN
GVTNLKNMASTTPDRQQACRCLQSAAKAVGPGLNTARAAGLPSACKVNIPYKISAS
TNCNTVR

SEQ ID NO:11, Nucleotide sequence of the open reading frame of Pk179

ATGGGGCTTGCTGTGGTGGACAAAAACACAGTTGCGATTTCTGCATCTGATGTTA
TGTTGTCCTTTGCTGCTTTTCCAGTCGAGATTCCTGGAGAGGTAGTATTTCTTCAT
CCCGTTCACAACTATGCTCTGATTGCGTATAATCCATCAGCAATGGATCCTGCCA
GTGCTTCAGTCATTCGTGCAGCTGAGCTACTACCTGAACCTGCACTCCAACGTGG
AGATTCAGTCTATCTTGTCGGATTGAGTAGGAACCTTCAAGCTACATCAAGAAAA
TCTATTGTAACCAATCCATGTGCAGCGTTAAACATTGGTTCTGCTGATTCTCCCCG
TTACAGAGCTACTAATATGGAAGTAATTGAGCTTGATACAGATTTTGGTAGCTCA
TTTTCAGGGGCGCTGACTGATGAGCAGGGAAGAATTCGGGCTATTTGGGGAAGT
TTTTCGACTCAGGTTAAATATAGTTCCACTTCTTCAGAAGACCACCAGTTTGTCAG
AGGTATCCCAGTATATGCAATCAGCCAAGTCCTTGAAAAAATCATAACCGGTGG
AAATGGACCAGCTCTTCTCATAAATGGTGTCAAAAGGCCAATGCCACTTGTTCGG
ATTTTGGAAGTTGAATTGTATCCTACTTTGCTTTCAAAAGCCCGGAGTTTTGGTCT
GAGTGATGAATGGATCCAAGTCCTAGTCAAGAAGGATCCTGTTAGACGTCAAGT
TCTGCGTGTTAAAGGTTGCCTGGCAGGATCAAAAGCTGAAAACCTTCTTGAACAA
GGCGATATGGTTCTGGCAGTCAATAAGATGCCAGTTACATGCTTCAATGACATAG
AAGCTGCTTGCCAAACATTGGATAAGGGTAGTTACAGCGATGAAAATCTCAATCT
AACAATCCTTAGACAGGGCCAAGAACTGGAGCTCGTAGTTGGAACTGATAAGAG
AGATGGGAATGGAACGACAAGAGTGATAAATTGGTGCGGATGCGTTGTTCAGGA
TCCTCATCCTGCGGTTCGTGCTCTTGGATTTCTTCCTGAGGAAGGTCATGGTGTCT
ATGTCACAAGATGGTGTCACGGGAGTCCCGCTCACCGATATGGCCTCTACGCGCT
TCAATGGATCGTGGAAGTTAATGGGAAGAAGACTCCTGACCTAAACGCATTCGC
AGATGCTACCAAGGAGCTAGAACACGGGCAGTTTGTGCGTATTAGGACTGTTCAT
CTAAACGGCAAGCCACGAGTATTGACCCTGAAACAAGATCTCCATTACTGGCCG
ACTTGGGAATTGAGGTTCGACCCAGAGACTGCTCTTTGGCGGAGAAATATATTGA
AAGCCTTGCAGTAA

SEQ ID NO:12, Deduced amino acid sequence of the open reading frame of Pk179

MGLAVVDKNTVAISASDVMLSFAAFPVEIPGEVVFLHPVHNYALIAYNPSAMDPASA
SVIRAAELLPEPALQRGDSVYLVGLSRNLQATSRKSIVTNPCAALNIGSADSPRYRAT
NMEVIELDTDFGSSFSGALTDEQGRIRAIWGSFSTQVKYSSTSSEDHQFVRGIPVYAIS
QVLEKIITGGNGPALLINGVKRPMPLVRILEVELYPTLLSKARSFGLSDEWIQVLVKK
DPVRRQVLRVKGCLAGSKAENLLEQGDMVLAVNKMPVTCFNDIEAACQTLDKGSY
SDENLNLTILRQGQELELVVGTDKRDGNGTTRVINWCGCVVQDPHPAVRALGFLPE
EGHGVYVTRWCHGSPAHRYGLYALQWIVEVNGKKTPDLNAFADATKELEHGQFVR
IRTVHLNGKPRVLTLKQDLHYWPTWELRFDPETALWRRNILKALQ

SEQ ID NO:13, Nucleotide sequence of the open reading frame of Pk202
ATGGCGTTCACGGCGCTTGTGTTCATTGTGTTCGTGGTGGGTGTCATGGTTTCTCC
AGTTTCAATCAGAGCAACTGAGGTCAAACTTTCTGGAGGAGAAGCTGATGTAAC
GTGTGATGCAGTACAGCTTAGTTCATGCGCAACACCAATGCTCACAGGAGTACCA
CCGTCTACAGAGTGTTGCGGGAAACTGAAGGAGCAACAGCCGTGTTTTTGTACAT
ATATTAAAGATCCAAGATATAGTCAATATGTTGGTTCTGCAAATGCTAAGAAAAC
GTTAGCAACTTGTGGTGTTCCTTATCCTACTTGTTGA

SEQ ID NO:14, Deduced amino acid sequence of the open reading frame of Pk202
MAFTALVFIVFVVGVMVSPVSIRATEVKLSGGEADVTCDAVQLSSCATPMLTGVPPS
TECCGKLKEQQPCFCTYIKDPRYSQYVGSANAKKTLATCGVPYPTC

SEQ ID NO:15, Nucleotide sequence of the open reading frame of Pk206
ATGGCCCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCTGAGGAAGGGCTTG
TTCGTCAGGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACAGTTGATGA
GGTCTGGAGAGATATCCAACAGGACAAGAATGGAAACGGTACTAGTACTACTAC
TACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGA
GCTGGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTT
CAAATGGGCAATGGGTTGAGTATCATCATCAGCCTCAACAACAACAAGGGTTTA
TGACATATCCGGTTTGCGAGATGCAAGATATGGTGATGATGGGTGGATTATCGGA
TACACCACAAGCGCCTGGGAGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGA
CTGTTGAGAGGAGACAGAAGAGGATGATCAAGAACAGAGAATCTGCAGCACGTT
CACGAGCTAGGAAACAGGCTTATACACATGAATTAGAGATCAAGGTTTCAAGGT
TAGAAGAAGAAAACGAAAAACTTCGGAGGCTAAAGGAGGTGGAGAAGATCCTA
CCAAGTGAACCACCACCAGATCCTAAGTGGAAGCTCCGGCGAACAAACTCTGCT
TCTCTCTGA

SEQ ID NO:16, Deduced amino acid sequence of the open reading frame of Pk206
MALDELLKTVLPPAEEGLVRQGSLTLPRDLSKKTVDEVWRDIQQDKNGNGTSTTTT
HKQPTLGEITLEDLLLRAGVVTETVVPQENVVNIASNGQWVEYHHQPQQQQGFMTY
PVCEMQDMVMMGGLSDTPQAPGRKRVAGEIVEKTVERRQKRMIKNRESAARSRAR
KQAYTHELEIKVSRLEEENEKLRRLKEVEKILPSEPPPDPKWKLRRTNSASL

SEQ ID NO:17, Nucleotide sequence of the open reading frame of Pk207
ATGGCGCAATCCCGATTATTAGCGTTTGCTTCAGCGGCGCGTTCACGTGTTCGAC
CAATCGCTCAAAGGCGTTTAGCGTTTGGATCATCCACGTCTGGTCGCACAGCTGA
TCCAGAGATCCATGCCGGTAACGATGGAGCCGATCCAGCTATCTATCCGAGAGA
CCCTGAAGGTATGGATGATGTTGCAAACCCTAAAACGGCGGCGGAAGAAATCGT
AGACGATACTCCCCGACCGAGTTTAGAAGAGCAACCGCTTGTACCGCCGAAATC
TCCACGCGCCACTGCGCACAAGCTAGAGAGTACTCCCGTTGGTCACCCGTCAGAA
CCTCATTTCCAACAGAAACGAAAAAACTCCACCGCTTCTCCGCCGTCGCTTGATT
CCGTGAGCTGTGCTGGTTTAGACGGTTCACCATGGCCGAGAGACGAAGGAGAAG
TGGAAGAGCAAAGGCGAAGAGAAGATGAAACAGAGAGTGACCAAGAGTTTTAC
AAACACCACAAAGCTTCTCCGTTATCGGAGATTGAATTCGCCGATACTCGGAAAC
CTATTACGCAAGCTACCGATGGAACTGCCTACCCAGCCGGGAAAGATGTGATCG
GATGGTTACCGGAGCAGCTAGACACGGCGGAAGAATCTTTGATGAAAGCAACAA
TGATATTCAAACGCAACGCAGAACGTGGCGATCCTGAAACGTTTCCTCATTCTAG
AATCTTAAGAGAAATGAGAGGCGAGTGGTTTTAA

SEQ ID NO:18, Deduced amino acid sequence of the open reading frame of Pk207
MAQSRLLAFASAARSRVRPIAQRRLAFGSSTSGRTADPEIHAGNDGADPAIYPRDPEG
MDDVANPKTAAEEIVDDTPRPSLEEQPLVPPKSPRATAHKLESTPVGHPSEPHFQQKR
KNSTASPPSLDSVSCAGLDGSPWPRDEGEVEEQRRREDETESDQEFYKHHKASPLSEI
EFADTRKPITQATDGTAYPAGKDVIGWLPEQLDTAEESLMKATMIFKRNAERGDPET
FPHSRILREMRGEWF

SEQ ID NO:19, Nucleotide sequence of the open reading frame of Pk209
ATGTCCGTGGCTCGATTCGATTTCTCTTGGTGCGATGCTGATTATCACCAGGAGA
CGCTGGAGAATCTGAAGATAGCTGTGAAGAGCACTAAGAAGCTTTGTGCTGTTAT
GCTAGACACTGTAGGACCTGAGTTGCAAGTTATTAACAAGACTGAGAAAGCTAT
TTCTCTTAAAGCTGATGGCCTTGTAACTTTGACTCCGAGTCAAGATCAAGAAGCC
TCCTCTGAAGTCCTTCCCATTAATTTTGATGGGTTAGCGAAGGCGGTTAAGAAAG
GAGACACTATCTTTGTTGGACAATACCTCTTCACTGGTAGTGAAACAACTTCAGT
TTGGCTTGAGGTTGAAGAAGTTAAAGGAGATGATGTCATTTGTATTTCAAGGAAT
GCTGCTACTCTGGGTGGTCCGTTATTCACATTGCACGTCTCTCAAGTTCACATTGA
TATGCCAACCCTAACTGAGAAGGATAAGGAGGTTATAAGTACATGGGGAGTTCA
GAATAAGATCGACTTTCTCTCATTATCTTATTGTCGACATGCAGAAGATGTTCGC
CAGGCCCGTGAGTTGCTTAACAGTTGTGGTGACCTCTCTCAAACACAAATATTTG
CGAAGATTGAGAATGAAGAGGGACTAACCCACTTTGACGAAATTCTACAAGAAG
CAGATGGCATTATTCTTTCTCGTGGGAATTTGGGTATCGATCTACCTCCGGAAAA
GGTGTTTTTGTTCCAAAAGGCTGCTCTTTACAAGTGTAACATGGCTGGAAAGCCT
GCCGTTCTTACTCGTGTTGTAGACAGTATGACAGACAATCTGCGGCCAACTCGTG
CAGAGGCAACTGATGTTGCTAATGCTGTTTTAGATGGAAGTGATGCAATTCTTCT
TGGTGCTGAGACTCTTCGTGGATTGTACCCTGTTGAAACCATATCAACTGTTGGT
AGAATCTGTTGTGAGGCAGAGAAAGTTTTCAACCAAGATTTGTTCTTTAAGAAGA
CTGTCAAGTATGTTGGAGAACCAATGACTCACTTGGAATCTATTGCTTCTTCTGCT
GTACGGGCAGCAATCAAGGTTAAGGCATCCGTAATTATATGCTTCACCTCGTCTG
GCAGAGCAGCAAGGTTGATTGCCAAATACCGTCCAACTATGCCCGTTCTCTCTGT
TGTCATTCCCCGACTTACGACAAATCAGCTGAAGTGGAGCTTTAGCGGAGCCTTT
GAGGCAAGGCAGTCACTTATTGTCAGAGGTCTTTTCCCCATGCTTGCTGATCCTC
GTCACCCTGCGGAATCAACAAGTGCAACAAATGAGTCGGTTCTTAAAGTGGCTCT
AGACCATGGGAAGCAAGCCGGAGTGATCAAGTCACATGACAGAGTTGTGGTCTG
TCAGAAAGTGGGAGATGCGTCCGTGGTCAAAATCATCGAGCTAGAGGATTAG

SEQ ID NO:20, Deduced amino acid sequence of the open reading frame of Pk209
MSVARFDFSWCDADYHQETLENLKIAVKSTKKLCAVMLDTVGPELQVINKTEKAIS
LKADGLVTLTPSQDQEASSEVLPINFDGLAKAVKKGDTIFVGQYLFTGSETTSVWLE
VEEVKGDDVICISRNAATLGGPLFTLHVSQVHIDMPTLTEKDKEVISTWGVQNKIDFL
SLSYCRHAEDVRQARELLNSCGDLSQTQIFAKIENEEGLTHFDEILQEADGIILSRGNL
GIDLPPEKVFLFQKAALYKCNMAGKPAVLTRVVDSMTDNLRPTRAEATDVANAVL
DGSDAILLGAETLRGLYPVETISTVGRICCEAEKVFNQDLFFKKTVKYVGEPMTHLES
IASSAVRAAIKVKASVIICFTSSGRAARLIAKYRPTMPVLSVVIPRLTTNQLKWSFSGA
FEARQSLIVRGLFPMLADPRHPAESTSATNESVLKVALDHGKQAGVIKSHDRVVVCQ
KVGDASVVKIIELED

SEQ ID NO:21, Nucleotide sequence of the open reading frame of Pk215
ATGGCGATTTACAGATCTCTAAGAAAGCTAGTTGAAATCAATCACCGGAAAACA
AGACCATTCCTCACCGCCGCTACAGCTTCCGGCGGAACCGTTTCTCTGACTCCAC
CGCAGTTTTCGCCGTTGTTCCCACATTTCTCACACCGTTTATCTCCGCTTTCGAAA

TGGTTCGTTCCTCTTAATGGACCTCTCTTCTTATCTTCTCCTCCTTGGAAACTTCTC
CAGTCTGCGACACCTTTGCACTGGCGCGGAAACGGCTCTGTTTTGAAAAAAGTCG
AAGCTCTGAATCTTAGATTGGATCGAATTAGAAGCAGAACTAGGTTTCCGAGAC
AGTTAGGGTTACAGTCTGTGGTACCAAACATATTGACGGTGGATCGCAACGATTC
CAAGGAAGAAGATGGTGGAAAATTAGTCAAGAGTTTTGTTAATGTGCCGAATAT
GATATCAATGGCGAGATTAGTATCTGGTCCTGTGCTTTGGTGGATGATCTCGAAT
GAGATGTATTCTTCTGCTTTCTTAGGGTTGGCTGTTTCTGGAGCTAGTGATTGGTT
AGATGGTTACGTGGCTCGGAGGATGAAGATTAACTCTGTGGTTGGCTCGTACCTT
GATCCTCTTGCAGACAAGGTTCTTATCGGGTGTGTAGCAGTAGCAATGGTGCAGA
AGGATCTCTTACATCCTGGACTGGTTGGAATTGTGTTGTTACGGGATGTTGCACT
CGTTGGTGGTGCAGTTTACCTAAGGGCACTAAACTTGGACTGGAGGTGGAAAAC
TTGGAGTGACTTCTTCAATCTAGATGGTTCAAGTCCTCAGAAAGTAGAACCATTG
TTTATAAGCAAGGTGAATACAGTTTTCCAGTTGACTCTAGTCGCTGGTGCAATAC
TTCAACCAGAGTTTGGGAATCCAGACACCCAGACATGGATCACTTATCTAAGGTA
A

SEQ ID NO:22, Deduced amino acid sequence of the open reading frame of Pk215
MAIYRSLRKLVEINHRKTRPFLTAATASGGTVSLTPPQFSPLFPHFSHRLSPLSKWFVP
LNGPLFLSSPPWKLLQSATPLHWRGNGSVLKKVEALNLRLDRIRSRTRFPRQLGLQS
VVPNILTVDRNDSKEEDGGKLVKSFVNVPNMISMARLVSGPVLWWMISNEMYSSAF
LGLAVSGASDWLDGYVARRMKINSVVGSYLDPLADKVLIGCVAVAMVQKDLLHPG
LVGIVLLRDVALVGGAVYLRALNLDWRWKTWSDFFNLDGSSPQKVEPLFISKVNTV
FQLTLVAGAILQPEFGNPDTQTWITYLR

SEQ ID NO:23, Nucleotide sequence of the open reading frame of Pk239
ATGGTAAAGGAAACTCTAATTCCTCCGTCATCTACGTCAATGACGACCGGAACAT
CTTCTTCTTCGTCTCTTTCAATGACGTTATCCTCAACAAACGCGTTATCGTTTTTGT
CGAAAGGATGGAGAGAGGTATGGGATTCAGCAGATGCGGATTTGCAGCTGATGC
GAGACAGAGCTAACTCTGTTAAGAATCTAGCATCAACGTTCGATAGAGAGATCG
AGAATTTCCTCAATAACTCGGCGAGGTCTGCGTTTCCCGTTGGTTCACCATCGGC
GTCGTCTTTCTCAAATGAAATTGGTATCATGAAGAAGCTTCAGCCGAAGATTTCG
GAGTTTCGTAGGGTTTATTCGGCGCCGGAGATTAGTCGCAAGGTTATGGAGAGAT
GGGGACCTGCGAGAGCGAAGCTTGGAATGGATCTATCGGCGATTAAGAAGGCGA
TTGTGTCTGAGATGGAATTGGATGAGCGTCAGGGAGTTTTGGAGATGAGTAGATT
GAGGAGACGGCGTAATAGTGATAGGGTTAGGTTTACGGAGTTTTTCGCGGAGGC
TGAGAGAGATGGAGAAGCTTATTTCGGTGATTGGGAACCGATTAGGTCTTTGAA
GAGTAGATTTAAAGAGTTTGAGAAACGAAGCTCGTTAGAAATATTGAGTGGATT
CAAGAACAGTGAATTTGTTGAGAAGCTCAAAACCAGCTTTAAATCAATTTACAA
AGAAACTGATGAGGCTAAGGATGTCCCTCCGTTGGATGTACCTGAACTGTTGGCA
TGTTTGGTTAGACAATCTGAACCTTTTCTTGATCAGATTGGTGTTAGAAAGGATA
CATGTGACCGAATAGTAGAAAGCCTTTGCAAATGCAAGAGCCAACAACTTTGGC
GTCTGCCATCTGCACAAGCATCCGATTTAATTGAAAATGATAACCATGGAGTTGA
TTTGGATATGAGGATAGCCAGTGTTCTTCAAAGCACAGGACACCATTATGATGGT
GGGTTTTGGACTGATTTTGTGAAGCCTGAGACACCGGAAAACAAAAGGCATGTG
GCAATTGTTACAACAGCTAGTCTTCCTTGGATGACCGGAACAGCTGTAAATCCGC
TATTCAGAGCGGCGTATTTGGCAAAAGCTGCAAAACAGAGTGTTACTCTCGTGGT
TCCTTGGCTCTGCGAATCTGATCAAGAACTAGTGTATCCAAACAATCTCACCTTC
AGCTCACCTGAAGAACAAGAGAGTTATATACGTAAATGGTTGGAGGAAAGGATT
GGTTTCAAGGCTGATTTTAAAATCTCCTTTTACCCAGGAAAGTTTTCAAAAGAAA
GGCGCAGCATATTTCCTGCTGGTGACACTTCTCAATTTATATCGTCAAAAGATGC

TGACATTGCTATACTTGAAGAACCTGAACATCTCAACTGGTATTATCACGGCAAG
CGTTGGACTGATAAATTCAACCATGTTGTTGGAATTGTCCACACAAACTACTTAG
AGTACATCAAGAGGGAGAAGAATGGAGCTCTTCAAGCATTTTTTGTGAACCATGT
AAACAATTGGGTCACACGAGCGTATTGTGACAAGGTTCTTCGCCTCTCTGCGGCA
ACACAAGATTTACCAAAGTCTGTTGTATGCAATGTCCATGGTGTCAATCCCAAGT
TCCTTATGATTGGGGAGAAAATTGCTGAAGAGAGATCCCGTGGTGAACAAGCTTT
CTCAAAAGGTGCATACTTCTTAGGAAAAATGGTGTGGGCTAAAGGATACAGAGA
ACTAATAGATCTGATGGCTAAACACAAAAGCGAACTTGGGAGCTTCAATCTAGA
TGTATATGGGAACGGTGAAGATGCAGTCGAGGTCCAACGTGCAGCAAAGAAACA
TGACTTGAATCTCAATTTCCTCAAAGGAAGGGACCACGCTGACGATGCTCTTCAC
AAGTACAAAGTGTTCATAAACCCCAGCATCAGCGATGTTCTATGCACAGCAACC
GCAGAAGCACTAGCCATGGGGAAGTTTGTGGTGTGTGCAGATCACCCTTCAAAC
GAATTCTTTAGATCATTCCCGAACTGCTTAACTTACAAAACATCCGAAGACTTTG
TGTCCAAAGTGCAAGAAGCAATGACGAAAGAGCCACTACCTCTCACTCCTGAAC
AAATGTACAATCTCTCTTGGGAAGCAGCAACACAGAGGTTCATGGAGTATTCAG
ATCTCGATAAGATCTTAAACAATGGAGAGGGAGGAAGGAAGATGCGAAAATCA
AGATCGGTTCCGAGCTTTAACGAGGTGGTCGATGGAGGATTGGCATTCTCACACT
ATGTTCTAACAGGGAACGATTTCTTGAGACTATGCACTGGAGCAACACCAAGAA
CAAAAGACTATGATAATCAACATTGCAAGGATCTGAATCTCGTACCACCTCACGT
TCACAAGCCAATCTTCGGCTGGTAG

SEQ ID NO:24, Deduced amino acid sequence of the open reading frame of Pk239
MVKETLIPPSSTSMTTGTSSSSSLSMTLSSTNALSFLSKGWREVWDSADADLQLMRD
RANSVKNLASTFDREIENFLNNSARSAFPVGSPSASSFSNEIGIMKKLQPKISEFRRVYS
APEISRKVMERWGPARAKLGMDLSAIKKAIVSEMELDERQGVLEMSRLRRRRNSDR
VRFTEFFAEAERDGEAYFGDWEPIRSLKSRFKEFEKRSSLEILSGFKNSEFVEKLKTSF
KSIYKETDEAKDVPPLDVPELLACLVRQSEPFLDQIGVRKDTCDRIVESLCKCKSQQL
WRLPSAQASDLIENDNHGVDLDMRIASVLQSTGHHYDGGFWTDFVKPETPENKRHV
AIVTTASLPWMTGTAVNPLFRAAYLAKAAKQSVTLVVPWLCESDQELVYPNNLTFS
SPEEQESYIRKWLEERIGFKADFKISFYPGKFSKERRSIFPAGDTSQFISSKDADIAILEE
PEHLNWYYHGKRWTDKFNHVVGIVHTNYLEYIKREKNGALQAFFVNHVNNWVTR
AYCDKVLRLSAATQDLPKSVVCNVHGVNPKFLMIGEKIAEERSRGEQAFSKGAYFL
GKMVWAKGYRELIDLMAKHKSELGSFNLDVYGNGEDAVEVQRAAKKHDLNLNFL
KGRDHADDALHKYKVFINPSISDVLCTATAEALAMGKFVVCADHPSNEFFRSFPNCL
TYKTSEDFVSKVQEAMTKEPLPLTPEQMYNLSWEAATQRFMEYSDLDKILNNGEGG
RKMRKSRSVPSFNEVVDGGLAFSHYVLTGNDFLRLCTGATPRTKDYDNQHCKDLNL
VPPHVHKPIFGW

SEQ ID NO:25, Nucleotide sequence of the open reading frame of Pk240
ATGGCGACTTTTGCTGAACTTGTTTTATCGACTTCTCGCTGTACATGCCCTTGCCG
TTCATTCACTAGAAAACCCCTAATTCGTCCCCCTTTATCTGGTCTGCGTCTCCCCG
GTGATACCAAACCATTGTTTCGTTCCGGACTTGGTCGGATTTCTGTTAGCCGGCGT
TTCCTCACGGCCGTTGCTCGAGCTGAATCAGACCAGCTTGGTGATGATGACCACT
CAAAGGGAATTGATAGAATCCATAACTTGCAGAATGTGGAAGATAAGCAGAAGA
AAGCAAGCCAGCTTAAGAAAAGAGTGATCTTTGGTATTGGCATTGGTTTACCTGT
TGGATGTGTTGTGTTAGCTGGAGGATGGGTTTTCACTGTAGCTTTAGCATCTTCTG
TTTTTATCGGTTCCCGCGAATATTTCGAGCTTGTTAGAAGTAGAGGCATAGCTAA
AGGAATGACTCCTCCTCCACGATATGTATCTCGAGTTTGCTCGGTTATATGTGCCC
TTATGCCCATACTTACACTGTACTTTGGTAACATTGATATATTGGTGACATCTGCA
GCATTTGTTGTTGCAATAGCATTGTTAGTACAAAGAGGATCCCCACGTTTTGCTC

AGCTGAGTAGTACAATGTTTGGTCTGTTTTACTGTGGTTATCTCCCTTCTTTCTGG
GTTAAGCTTCGCTGTGGTTTAGCTGCTCCTGCGCTTAACACTGGTATCGGAAGGA
CATGGCCAATTCTTCTTGGTGGTCAAGCTCATTGGACAGTTGGACTTGTGGCAAC
ATTGATTTCTTTCAGCGGTGTAATTGCGACAGACACATTTGCTTTTCTCGGTGGAA
AGACTTTTGGTAGGACACCTCTTACTAGTATTAGTCCCAAGAAGACATGGGAAGG
AACTATTGTAGGACTTGTTGGTTGTATAGCCATTACCATATTACTCTCTAAATATC
TCAGTTGGCCACAATCTCTGTTCAGCTCAGTAGCTTTTGGGTTTCTTAACTTCTTT
GGGTCAGTCTTTGGTGATCTTACTGAATCAATGATCAAGCGTGATGCTGGCGTCA
AAGACTCTGGTTCACTTATCCCAGGACACGGTGGAATATTAGATAGAGTTGATAG
TTACATTTTCACCGGCGCATTAGCTTATTCATTCATCAAAACATCCCTAAAACTTT
ACGGAGTTTGA

SEQ ID NO:26, Deduced amino acid sequence of the open reading frame of Pk240
MATFAELVLSTSRCTCPCRSFTRKPLIRPPLSGLRLPGDTKPLFRSGLGRISVSRRFLTA
VARAESDQLGDDDHSKGIDRIHNLQNVEDKQKKASQLKKRVIFGIGIGLPVGCVVLA
GGWVFTVALASSVFIGSREYFELVRSRGIAKGMTPPPRYVSRVCSVICALMPILTLYF
GNIDILVTSAAFVVAIALLVQRGSPRFAQLSSTMFGLFYCGYLPSFWVKLRCGLAAPA
LNTGIGRTWPILLGGQAHWTVGLVATLISFSGVIATDTFAFLGGKTFGRTPLTSISPKK
TWEGTIVGLVGCIAITILLSKYLSWPQSLFSSVAFGFLNFFGSVFGDLTESMIKRDAGV
KDSGSLIPGHGGILDRVDSYIFTGALAYSFIKTSLKLYGV

SEQ ID NO:27, Nucleotide sequence of the open reading frame of Pk241
ATGGCTCAAACCATGCTGCTTACTTCAGGCGTCACCGCCGGCCATTTTTTGAGGA
ACAAGAGCCCTTTGGCTCAGCCCAAAGTTCACCATCTCTTCCTCTCTGGAAACTC
TCCGGTTGCACTACCATCTAGGAGACAATCATTCGTTCCTCTCGCTCTCTTCAAAC
CCAAAACCAAAGCTGCTCCTAAAAAGGTTGAGAAGCCGAAGAGCAAGGTTGAGG
ATGGCATCTTTGGAACGTCTGGTGGGATTGGTTTCACAAAGGCGAATGAGCTATT
CGTTGGTCGTGTTGCTATGATCGGTTTCGCTGCATCGTTGCTTGGTGAGGCGTTGA
CGGGAAAAGGGATATTAGCTCAGCTGAATCTGGAGACAGGGATACCGATTTACG
AAGCAGAGCCATTGCTTCTCTTCTTCATCTTGTTCACTCTGTTGGGAGCCATTGGA
GCTCTCGGAGACAGAGGAAAATTCGTCGACGATCCTCCCACCGGGCTCGAGAAA
GCCGTCATTCCTCCCGGCAAAAACGTCCGATCTGCCCTCGGTCTCAAAGAACAAG
GTCCATTGTTTGGGTTCACGAAGGCGAACGAGTTATTCGTAGGAAGATTGGCACA
GTTGGGAATAGCATTTTCACTGATAGGAGAGATTATTACCGGGAAAGGAGCATT
AGCTCAACTCAACATTGAGACCGGTATACCAATTCAAGATATCGAACCACTTGTC
CTCTTAAACGTTGCTTTCTTCTTCTTCGCTGCCATTAATCCTGGTAATGGAAAATT
CATCACCGATGATGGTGAAGAAAGCTAA

SEQ ID NO:28, Deduced amino acid sequence of the open reading frame of Pk241
MAQTMLLTSGVTAGHFLRNKSPLAQPKVHHLFLSGNSPVALPSRRQSFVPLALFKPK
TKAAPKKVEKPKSKVEDGIFGTSGGIGFTKANELFVGRVAMIGFAASLLGEALTGKGI
LAQLNLETGIPIYEAEPLLLFFILFTLLGAIGALGDRGKFVDDPPTGLEKAVIPPGKNVR
SALGLKEQGPLFGFTKANELFVGRLAQLGIAFSLIGEIITGKGALAQLNIETGIPIQDIEP
LVLLNVAFFFFAAINPGNGKFITDDGEES

SEQ ID NO:29, Nucleotide sequence of the open reading frame of Pk242
ATGGGTGCAGGTGGAAGAATGCCGGTTCCTACTTCTTCCAAGAAATCGGAAACC
GACACCACAAAGCGTGTGCCGTGCGAGAAACCGCCTTTCTCGGTGGGAGATCTG
AAGAAAGCAATCCCGCCGCATTGTTTCAAACGCTCAATCCCTCGCTCTTTCTCCT
ACCTTATCAGTGACATCATTATAGCCTCATGCTTCTACTACGTCGCCACCAATTAC

TTCTCTCTCCTCCCTCAGCCTCTCTCTTACTTGGCTTGGCCACTCTATTGGGCCTGT
CAAGGCTGTGTCCTAACTGGTATCTGGGTCATAGCCCACGAATGCGGTCACCACG
CATTCAGCGACTACCAATGGCTGGATGACACAGTTGGTCTTATCTTCCATTCCTTC
CTCCTCGTCCCTTACTTCTCCTGGAAGTATAGTCATCGCCGTCACCATTCCAACAC
TGGATCCCTCGAAAGAGATGAAGTATTTGTCCCAAAGCAGAAATCAGCAATCAA
GTGGTACGGGAAATACCTCAACAACCCTCTTGGACGCATCATGATGTTAACCGTC
CAGTTTGTCCTCGGGTGGCCCTTGTACTTAGCCTTTAACGTCTCTGGCAGACCGTA
TGACGGGTTCGCTTGCCATTTCTTCCCCAACGCTCCCATCTACAATGACCGAGAA
CGCCTCCAGATATACCTCTCTGATGCGGGTATTCTAGCCGTCTGTTTTGGTCTTTA
CCGTTACGCTGCTGCACAAGGGATGGCCTCGATGATCTGCCTCTACGGAGTACCG
CTTCTGATAGTGAATGCGTTCCTCGTCTTGATCACTTACTTGCAGCACACTCATCC
CTCGTTGCCTCACTACGATTCATCAGAGTGGGACTGGCTCAGGGGAGCTTTGGCT
ACCGTAGACAGAGACTACGGAATCTTGAACAAGGTGTTCCACAACATTACAGAC
ACACACGTGGCTCATCACCTGTTCTCGACAATGCCGCCTTATAACGCAATGGAAG
CTACAAAGGCGATAAAGCCAATTCTGGGAGACTATTACCAGTTCGATGGAACAC
CGTGGTATGTAGCGATGTATAGGGAGGCAAAGGAGTGTATCTATGTAGAACCGG
ACAGGGAAGGTGACAAGAAAGGTGTGTACTGGTACAACAATAAGTTATGA

SEQ ID NO:30, Deduced amino acid sequence of the open reading frame of Pk242

MGAGGRMPVPTSSKKSETDTTKRVPCEKPPFSVGDLKKAIPPHCFKRSIPRSFSYLISD
IIIASCFYYVATNYFSLLPQPLSYLAWPLYWACQGCVLTGIWVIAHECGHHAFSDYQ
WLDDTVGLIFHSFLLVPYFSWKYSHRRHHSNTGSLERDEVFVPKQKSAIKWYGKYL
NNPLGRIMMLTVQFVLGWPLYLAFNVSGRPYDGFACHFFPNAPIYNDRERLQIYLSD
AGILAVCFGLYRYAAAQGMASMICLYGVPLLIVNAFLVLITYLQHTHPSLPHYDSSE
WDWLRGALATVDRDYGILNKVFHNITDTHVAHHLFSTMPPYNAMEATKAIKPILGD
YYQFDGTPWYVAMYREAKECIYVEPDREGDKKGVYWYNNKL

SEQ ID NO:31, Nucleotide sequence of the open reading frame of Bn011

ATGGCTTCAATAAATGAAGATGTGTCTATTGGAAACTTAGGCAGTCTCCAAACAC
TCCCAGACTCATTCACCTGGAAACTCACCGCTGCTGACTCCATTCTCCCTCCCTCC
TCCGCCGCTGTGAAAGAGTCCATTCCGGTCATCGACCTCTCCGATCCTGACGTCA
CCAATTTGTTAGGAAATGCATGCAAAACGTGGGGAGCGTTTCAGATAGCCAACC
ACGGGGTCTCTCAAAGTCTCCTCGACGACGTTGAATCTCTCTCCAAAACCTTTTTC
GATATGCCGTCAGAGAGGAAACTCGAGGCTGCTTCCTCTAATAAAGGAGTTAGT
GGGTACGGAGAACCTCGAATCTCTCTTTTCTTCGAGAAGAAAATGTGGTCTGAAG
GGTTGACAATCGCCGACGGCTCCTACCGCAACCAGTTCCTTACTATTTGGCCCCG
TGATTACACCAAATACTGCGGAATAATCGAAGAGTACAAGGGTGAAATGGAAAA
ATTAGCAAGCAGACTTCTATCATGCATATTAGGATCACTTGGTGTCACCGTAGAC
GACATCGAATGGGCTAAGAAGACCGAGAAATCTGAATCAAAAATGGGCCAAAG
CGTCATACGACTAAACCATTACCCGGTTTGTCCTGAGCCAGAAAGAGCCATGGGT
CTAGCCGCTCATACCGACTCATGTCTTCTAACCATTTTGCACCAGAGCAACATGG
GAGGGCTACAAGTGTTCAAAGAAGAGTCCGGTTGGGTTACGGTAGAGCCCATTC
CTGGTGTTCTTGTGGTCAACATCGGCGACCTCTTTCACATTCTATCGAATGGGAA
GTTTCCTAGCGTGGTTCACCGAGCAAGGGTTAACCGAACCAAGTCAAGAATATC
GATAGCGTATCTGTGGGGTGGTCCAGCCGGTGAAGTGGAGATAAGTCCAATATC
AAAGATAGTTGGTCCGGTTGGACCGTGTCTATACCGGCCAGTTACTTGGAGTGAA
TATCTCCGAATCAAATTTGAGGTTTTCGACAAGGCATTGGACGCAATTGGAGTCG
TTAATCCCACCAATTGA

SEQ ID NO:32, Deduced amino acid sequence of the open reading frame of Bn011

MASINEDVSIGNLGSLQTLPDSFTWKLTAADSILPPSSAAVKESIPVIDLSDPDVTNLL
GNACKTWGAFQIANHGVSQSLLDDVESLSKTFFDMPSERKLEAASSNKGVSGYGEP
RISLFFEKKMWSEGLTIADGSYRNQFLTIWPRDYTKYCGIIEEYKGEMEKLASRLLSCI
LGSLGVTVDDIEWAKKTEKSESKMGQSVIRLNHYPVCPEPERAMGLAAHTDSCLLTI
LHQSNMGGLQVFKEESGWVTVEPIPGVLVVNIGDLFHILSNGKFPSVVHRARVNRTK
SRISIAYLWGGPAGEVEISPISKIVGPVGPCLYRPVTWSEYLRIKFEVFDKALDAIGVV
NPTN

SEQ ID NO:33, Nucleotide sequence of the open reading frame of Bn077
ATGGCTACATTCTCTTGTAATTCTTATGAACAAAATCACGCTCCTTTCGACCGTCA
CGCTAATGATACTGATATTGATGATCCTGATCATGATCATCATGATGGTGTTCAG
CAAGAGGAGAGTGGATGGACAACTTATCTTGAAGATTTCTCAAATCAATACAGA
ACTCATCCTGAAGATAACGATCATCAAGATAAGAGTTCGTGTTCGATTCTGGACG
CCTCTCCTTCTCTGGTCTCCGACGCCGCCACTGACGCATTTTCTGGCCGGAGTTTT
CCAGTTAATTTTCCGGTGAAATTGAAGTTTGGGAAGGCAAGAACCAAAAAGATT
TGTGAGGATGATTCTTTGGAGGATACGGCTAGCTCTCCGGTTAATAGCCCTAAGG
TCAGTCAGATTGAACATATTCAGACGCCTCCTAGAAAACATGAGGACTATGTCTC
TTCTAGTTTCGTTATGGGAAATATGAGTGGCATGGGGGGATCATCAAATCCAAATC
CAAGAAGGAGATGAACAAAAGTTGACGATGATGAGGAATCTCAGAGAAGGAAA
CAACAGTAACAGTAATAATATGGACTTGAGGGCTAGAGGATTATGCGTCGTCCCT
ATTTCCATGTTGGGTAATTTTAATGGCCGCTTCTGA

SEQ ID NO:34, Deduced amino acid sequence of the open reading frame of Bn077
MATFSCNSYEQNHAPFDRHANDTDIDDPDHDHHDGVQQEESGWTTYLEDFSNQYR
THPEDNDHQDKSSCSILDASPSLVSDAATDAFSGRSFPVNFPVKLKFGKARTKKICED
DSLEDTASSPVNSPKVSQIEHIQTPPRKHEDYVSSSFVMGNMSGMGDHQIQIQEGDEQ
KLTMMRNLREGNNSNSNNMDLRARGLCVVPISMLGNFNGRF

SEQ ID NO:35, Nucleotide sequence of the open reading frame of Jb001
ATGGCAACGGAATGCATTGCAACGGTCCCTCAAATATTCAGTGAAAACAAAACC
AAAGAGGATTCTTCGATCTTCGATGCAAAGCTCCTTAATCAGCACTCACACCACA
TACCTCAACAGTTCGTATGGCCCGACCACGAGAAACCTTCTACGGATGTTCAACC
TCTCCAAGTCCCACTCATAGACCTAGCCGGTTTCCTCTCCGGCGACTCGTGCTTGG
CATCGGAGGCTACTAGACTCGTCTCAAAGGCTGCAACGAAACATGGCTTCTTCCT
AATCACTAACCATGGTATCGATGAGAGCCTCTTGTCTCGTGCCTATCTGCATATG
GACTCTTTCTTTAAGGCCCCGGCTTGTGAGAAGCAGAAGGCTCAGAGGAAGTGG
GGTGAGAGCTCCGGTTACGCTAGTAGTTTCGTCGGGAGATTCTCCTCAAAGCTCC
CGTGGAAGGAGACTCTGTCGTTTAAGTTCTCTCCCGAGGAGAAGATCCATTCCCA
AACCGTTAAAGACTTTGTTTCTAAGAAAATGTGCGATGGATACGAAGATTTCGGG
AAGGTTTATCAAGAATACGCGGAGGCCATGAACACTCTCTCACTAAAGATCATG
GAGCTTCTTGGAATGAGTCTTGGGGTCGAGAGGAGATATTTTAAAGAGTTTTTCG
AAGACAGCGATTCAATATTCCGGTTGAATTACTACCCGCAGTGCAAGCAACCGG
AGCTTGCACTAGGGACAGGACCCCACTGCGACCCAACATCTCTAACCATACTTCA
TCAAGACCAAGTTGGCGGTCTGCAAGTTTTCGTGGACAACAAATGGCAATCCATT
CCTCCTAACCCTCACGCTTTCGTGGTGAACATAGGCGACACCTTCATGGCTCTAA
CGAATGGAAGATACAAGAGTTGTTTGCATCGGGCGGTGGTGAACAGCGAGAGAG
AAAGGAAGACGTTTGCATTCTTCCTATGTCCGAAAGGGGAAAAAGTGGTGAAGC
CACCAGAAGAACTAGTAAACGGAGTGAAGTCTGGTGAAAGAAAGTATCCTGATT
TTACGTGGTCTATGTTTCTCGAGTTCACACAGAAGCATTATAGGGCAGACATGAA
CACTCTTGACGAGTTCTCAATTTGGCTTAAGAACAGAAGAAGTTTCTAA

SEQ ID NO:36, Deduced amino acid sequence of the open reading frame of Jb001

MATECIATVPQIFSENKTKEDSSIFDAKLLNQHSHHIPQQFVWPDHEKPSTDVQPLQV
PLIDLAGFLSGDSCLASEATRLVSKAATKHGFFLITNHGIDESLLSRAYLHMDSFFKAP
ACEKQKAQRKWGESSGYASSFVGRFSSKLPWKETLSFKFSPEEKIHSQTVKDFVSKK
MCDGYEDFGKVYQEYAEAMNTLSLKIMELLGMSLGVERRYFKEFFEDSDSIFRLNY
YPQCKQPELALGTGPHCDPTSLTILHQDQVGGLQVFVDNKWQSIPPNPHAFVVNIGD
TFMALTNGRYKSCLHRAVVNSERERKTFAFFLCPKGEKVVKPPEELVNGVKSGERK
YPDFTWSMFLEFTQKHYRADMNTLDEFSIWLKNRRSF

SEQ ID NO:37, Nucleotide sequence of the open reading frame of Jb002

ATGGCGTCAGAGCAAGCAAGGAGAGAAAACAAGGTGACGGAGAGAGAAGTTCA
GGTGGAGAAAGACAGAGTCCCAAAGATGACGAGTCATTTCGAGTCCATGGCCGA
AAAAGGCAAAGATTCCGACACACACAGGCATCAAACAGAAGGTGGTGGGACAC
AGTTCGTGTCTCTCTCAGACAAGGGGAGTAACATGCCGGTTTCTGATGAAGGAGA
GGGAGAGACGAAGATGAAGAGGACTCAGATGCCTCACTCCGTTGGAAAATTCGT
TACTAGCAGCGATTCAGGAACAGGGAAGAAGAAGGATGAGAAAGAGGAGCATG
AGAAGGCGTCGCTAGAGGATATTCATGGGTATAGAGCCAATGCTCAGCAGAAGT
CAATGGATAGTATAAAAGCAGCAGAGGAAAGGTATAACAAGGCTAAGGAGAGT
TTGAGCCATAGTGGACAAGAAGCTCGTGGAGGAAGAGGTGAAGAAATGGTGGG
AAAAGGGCGGGACAGTGGTGTCCGTGTTTCTCACGTTGGGGCTGTTGGTGGCGGT
GGTGGAGGTGAGGAAAAAGAGAGTGGTGTACATGGCTTTCATGGGGAGAAAGC
ACGACATGCTGAGCTTTTGGCTGCCGGAGGTGAGGAGATGAGAGAACGTGAAGG
TAAAGAATCAGCAGGTGGTGTTGGTGGTCGTAGCGTAAAAGATACGGTAGCCGA
GAAAGGACAGCAAGCTAAGGAAAGTGTAGGAGAAGGTGCTCAGAAAGCGGGCA
GTGCTACGAGTGAGAAAGCTCAGAGAGCTTCCGAGTATGCAACAGAGAAAGGAA
AAGAAGCTGGAAATATGACAGCTGAACAGGCGGCGAGAGCAAAAGACTATGCT
CTGCAGAAAGCTGTTGAAGCTAAAGAGACTGCGGCGGAGAAAGCTCAGAGAGCT
TCCGAGTATATGAAGGAAACAGGAAGCACAGCGGCTGAACAGGCTGCGAGAGCT
AAAGATTACACTCTTCAGAAAGCTGTGGAAGCTAAAGATGTTGCAGCTGAGAAA
GCTCAGAGAGCTTCAGAATACATGACAGAGACAGGAAAACAAGCCGGAAATGTT
GCAGCTCAGAAAGGGCAAGAGGCAGCTTCAATGACAGCAAAAGCTAAAGATTAT
ACTGTTCAGAAAGCCGGTGAAGCAGCTGGGTACATAAAAGAAACGACAGTGGAA
GGAGGAAAAGGAGCTGCACATTATGCAGGAGTGGCAGCTGAGAAAGCCGCTGC
GGTTGGGTGGACAGCGGCACATTTCACCACGGAGAAAGTGGTGCAAGGGACGAA
AGCGGTTGCAGGTACAGTGGAAGGTGCTGTGGGGTACGCAGGGCATAAGGCGGT
GGAAGTAGGATCTAAGGCAGTGGACTTGACTAAGGAGAAAGCTGCAGTGGCTGC
TGATACGGTGGTTGGGTATACGGCGAGGAAGAAAGAGGAAGCTCAACACAGAG
ACCAAGAGATGCATCAGGGAGGTGAGGAAGAAAAGCAACCAGGGTTTGTCTCAG
GAGCAAGGAGAGACTTTGGAGAAGAGTACGGGGAAGAAAGAGGGAGTGAGAAA
GATGTCTACGGCTATGGAGCAAAAGGAATACCCGGAGAAGGGAGGGGAGATGTT
GGGGAGGCAGAGTACGGAAGAGGGAGTGAGAAAGATGTCTTCGGATATGGACC
AAAAGGCACGGTCGAAGAAGCAAGGAGAGACGTTGGAGAAGAATACGGAGGAG
GAAGAGGCAGTGAGAGATATGTTGAAGAAGAAGGGGTTGGAGCGGGAGGGGTG
CTTGGGGCAATCGGCGAGACTATAGCTGAGATTGCACAGACGACAAAGAACATA
GTGATTGGTGATGCGCCTGTGAGGACACATGAGCATGGAACTACTGATCCTGACT
ATATGAGACGGGAACATGGACAACGTTGA

SEQ ID NO:38, Amino acid sequence of the open reading frame of Jb002

MASEQARRENKVTEREVQVEKDRVPKMTSHFESMAEKGKDSDTHRHQTEGGGTQF
VSLSDKGSNMPVSDEGEGETKMKRTQMPHSVGKFVTSSDSGTGKKKDEKEEHEKAS
LEDIHGYRANAQQKSMDSIKAAEERYNKAKESLSHSGQEARGGRGEEMVGKGRDS
GVRVSHVGAVGGGGGGEEKESGVHGFHGEKARHAELLAAGGEEMREREGKESAG
GVGGRSVKDTVAEKGQQAKESVGEGAQKAGSATSEKAQRASEYATEKGKEAGNM
TAEQAARAKDYALQKAVEAKETAAEKAQRASEYMKETGSTAAEQAARAKDYTLQ
KAVEAKDVAAEKAQRASEYMTETGKQAGNVAAQKGQEAASMTAKAKDYTVQKA
GEAAGYIKETTVEGGKGAAHYAGVAAEKAAAVGWTAAHFTTEKVVQGTKAVAGT
VEGAVGYAGHKAVEVGSKAVDLTKEKAAVAADTVVGYTARKKEEAQHRDQEMH
QGGEEEKQPGFVSGARRDFGEEYGEERGSEKDVYGYGAKGIPGEGRGDVGEAEYGR
GSEKDVFGYGPKGTVEEARRDVGEEYGGGRGSERYVEEEGVGAGGVLGAIGETIAEI
AQTTKNIVIGDAPVRTHEHGTTDPDYMRREHGQR

SEQ ID NO:39, Nucleotide sequence of the open reading frame of Jb003

ATGGCTAAGTCTTGCTATTTCAGACCAGCTCTTCTTCTTCTGTTAGTTCTTTTGGTT
CATGCCGAGTCACGCGGTCGGTTCGAGCCAAAGATTCTTATGCCGACAGAGGAA
GCTAACCCGGCTGACCAAGACGGAGATGGTGTCGGTACAAGATGGGCGGTTCTC
GTCGCTGGTTCTTCTGGATATGGAAACTACAGACACCAGGCTGACATGTGTCACG
CATATCAAATACTAAGAAAAGGAGGTTTAAAGGAAGAGAACATAGTCGTTTTGA
TGTATGATGATATCGCAAACCACCCACTTAATCCTCGTCCGGGTACTCTCATCAA
CCATCCTGACGGTGACGATGTTTACGCCGGAGTCCCTAAGGACTATACTGGTAGT
AGTGTTACGGCTGCAAACTTCTACGCTGTACTCCTAGGCGACCAGAAGGCTGTTA
AAGGTGGAAGCGGTAAGGTCATCGCTAGCAAGCCCAACGATCACATTTTCGTAT
ATTATGCGGATCATGGTGGTCCCGGAGTTCTTGGGATGCCAAATACGCCTCACAT
ATATGCAGCTGATTTTATTGAAACGCTTAAGAAGAAGCATGCTTCCGGAACATAC
AAAGAGATGGTTATATACGTAGAAGCGTGTGAAAGTGGGAGTATTTTCGAAGGG
ATAATGCCAAAGGACTTGAACATTTACGTAACAACGGCTTCAAATGCACAAGAG
AGTAGTTATGGAACATATTGTCCTGGCATGAATCCGTCACCCCCATCTGAATATA
TCACTTGCTTAGGGGATTTATATAGTGTTGCTTGGATGGAAGATAGTGAGACTCA
CAATTTAAAGAAAGAGACCATAAAGCAACAATACCACACGGTGAAGATGAGGA
CATCAAACTACAATACCTACTCAGGTGGCTCTCATGTGATGGAATACGGTAACAA
TAGTATTAAGTCGGAGAAGCTTTATCTTTACCAAGGGTTTGATCCAGCCACCGTT
AATCTCCCACTAAACGAATTACCGGTCAAGTCAAAAATAGGAGTCGTTAACCAA
CGCGATGCGGACCTTCTCTTCCTTTGGCATATGTATCGGACATCGGAAGATGGGT
CAAGGAAGAAGGATGACACATTGAAGGAATTAACTGAGACAACAAGGCATAGG
AAACATTTAGATGCAAGCGTCGAATTGATAGCCACAATTTTGTTTGGTCCGACGA
TGAATGTTCTTAACTTGGTTAGAGAACCCGGTTTGCCTTTGGTTGACGATTGGGA
ATGTCTTAAATCGATGGTACGTGTATTTGAAGAGCATTGTGGATCACTAACGCAA
TATGGGATGAAACATATGCGAGCGTTTGCAAACGTTTGTAACAACGGTGTGTCCA
AAGAGCTGATGGAGGAAGCTTCTACTGCGGCATGCGGTGGTTATAGTGAGGCTC
GCTACACGGTGCATCCATCAATCTTAGGCTATAGCGCCTGA

SEQ ID NO:40, Deduced amino acid sequence of the open reading frame of Jb003

MAKSCYFRPALLLLLVLLVHAESRGRFEPKILMPTEEANPADQDGDGVGTRWAVLV
AGSSGYGNYRHQADMCHAYQILRKGGLKEENIVVLMYDDIANHPLNPRPGTLINHP
DGDDVYAGVPKDYTGSSVTAANFYAVLLGDQKAVKGGSGKVIASKPNDHIFVYYA
DHGGPGVLGMPNTPHIYAADFIETLKKKHASGTYKEMVIYVEACESGSIFEGIMPKD
LNIYVTTASNAQESSYGTYCPGMNPSPPSEYITCLGDLYSVAWMEDSETHNLKKETI
KQQYHTVKMRTSNYNTYSGGSHVMEYGNNSIKSEKLYLYQGFDPATVNLPLNELPV
KSKIGVVNQRDADLLFLWHMYRTSEDGSRKKDDTLKELTETTRHRKHLDASVELIA

TILFGPTMNVLNLVREPGLPLVDDWECLKSMVRVFEEHCGSLTQYGMKHMRAFAN
VCNNGVSKELMEEASTAACGGYSEARYTVHPSILGYSA

SEQ ID NO:41, Nucleotide sequence of the open reading frame of Jb005

ATGGACGGTGCCGGAGAATCACGACTCGGTGGTGATGGTGGTGGTGATGGTTCT
GTTGGAGTTCAGATCCGACAAACACGGATGCTACCGGATTTTCTCCAGAGCGTGA
ATCTCAAGTATGTGAAATTAGGTTACCATTACTTAATCTCAAATCTCTTGACTCTC
TGTTTATTCCCTCTCGCCGTTGTTATCTCCGTCGAAGCCTCTCAGATGAACCCAGA
TGATCTCAAACAGCTCTGGATCCATCTACAATACAATCTGGTTAGTATCATCATC
TGTTCAGCGATTCTAGTCTTCGGGTTAACGGTTTATGTTATGACCCGACCTAGACC
CGTTTACTTGGTTGATTTCTCTTGTTATCTCCCACCTGATCATCTCAAAGCTCCTTA
CGCTCGGTTCATGGAACATTCTAGACTCACCGGAGATTTCGATGACTCTGCTCTC
GAGTTTCAACGCAAGATCCTTGAGCGTTCTGGTTTAGGGGAAGACACTTATGTCC
CTGAAGCTATGCATTATGTTCCACCGAGAATTTCAATGGCTGCTGCTAGAGAAGA
AGCTGAACAAGTCATGTTTGGTGCTTTAGATAACCTTTTCGCTAACACTAATGTG
AAACCAAAGGATATTGGAATCCTTGTTGTGAATTGTAGTCTCTTTAATCCAACTC
CTTCGTTATCTGCAATGATTGTGAACAAGTATAAGCTTAGAGGTAACATTAGAAG
CTACAATCTAGGCGGTATGGGTTGCAGCGCGGGAGTTATCGCTGTGGATCTTGCT
AAAGACATGTTGTTGGTACATAGGAACACTTATGCGGTTGTTGTTTCTACTGAGA
ACATTACTCAGAATTGGTATTTTGGTAACAAGAAATCGATGTTGATACCGAACTG
CTTGTTTCGAGTTGGTGGCTCTGCGGTTTTGCTATCGAACAAGTCGAGGGACAAG
AGACGGTCTAAGTACAGGCTTGTACATGTAGTCAGGACTCACCGTGGAGCAGAT
GATAAAGCTTTCCGTTGTGTTTATCAAGAGCAGGATGATACAGGGAGAACCGGG
GTTTCGTTGTCGAAAGATCTAATGGCGATTGCAGGGGAAACTCTCAAAACCAATA
TCACTACATTGGGTCCTCTTGTTCTACCGATAAGTGAGCAGATTCTCTTCTTTATG
ACTCTAGTTGTGAAGAAGCTCTTTAACGGTAAAGTGAAACCGTATATCCCGGATT
TCAAACTTGCTTTCGAGCATTTCTGTATCCATGCTGGTGGAAGAGCTGTGATCGA
TGAGTTAGAGAAGAATCTGCAGCTTTCACCAGTTCATGTCGAGGCTTCGAGGATG
ACTCTTCATCGATTTGGTAACACATCTTCGAGCTCCATTTGGTATGAATTGGCTTA
CATTGAAGCGAAGGGAAGGATGCGAAGAGGTAATCGTGTTTGGCAAATCGCGTT
CGGAAGTGGATTTAAATGTAATAGCGCGATTTGGGAAGCATTAAGGCATGTGAA
ACCTTCGAACAACAGTCCTTGGGAAGATTGTATTGACAAGTATCCGGTAACTTTA
AGTTATTAG

SEQ ID NO:42, Deduced amino acid sequence of the open reading frame of Jb005

MDGAGESRLGGDGGGDGSVGVQIRQTRMLPDFLQSVNLKYVKLGYHYLISNLLTLC
LFPLAVVISVEASQMNPDDLKQLWIHLQYNLVSIIICSAILVFGLTVYVMTRPRPVYL
VDFSCYLPPDHLKAPYARFMEHSRLTGDFDDSALEFQRKILERSGLGEDTYVPEAMH
YVPPRISMAAAREEAEQVMFGALDNLFANTNVKPKDIGILVVNCSLFNPTPSLSAMIV
NKYKLRGNIRSYNLGGMGCSAGVIAVDLAKDMLLVHRNTYAVVVSTENITQNWYF
GNKKSMLIPNCLFRVGGSAVLLSNKSRDKRRSKYRLVHVVRTHRGADDKAFRCVY
QEQDDTGRTGVSLSKDLMAIAGETLKTNITTLGPLVLPISEQILFFMTLVVKKLFNGK
VKPYIPDFKLAFEHFCIHAGGRAVIDELEKNLQLSPVHVEASRMTLHRFGNTSSSSIW
YELAYIEAKGRMRRGNRVWQIAFGSGFKCNSAIWEALRHVKPSNNSPWEDCIDKYP
VTLSY

SEQ ID NO:43, Nucleotide sequence of the open reading frame of Jb007

ATGTCGAGAGCTTTGTCAGTCGTTTGTGTCTTGCTCGCCATATCCTTCGTCTGTGC
ACGTGCTCGTCAGGTGCCGGGAGAGTCTGATGAGGGAAAGACGACGGGACATGA
CGATACAACAACAATGCCCATGCATGCAAAAGCAGCTGATCAGTTACCACCAAA

GAGCGTCGGCGACAAAAAATGCATCGGAGGAGTTGCTGGAGTCGGTGGATTCGC
CGGAGTTGGTGGTGTTGCCGGCGTGGGAGGTCTAGGGATGCCACTCATCGGTGGT
CTTGGCGGGATCGGTAAGTATGGTGGCATAGGCGGTGCAGCTGGAATCGGTGGA
TTTCATAGTATAGGCGGTGTTGGCGGTCTAGGCGGTGTCGGAGGAGGTGTTGGCG
GTCTAGGCGGTGTTGGAGGGGGTGTTGGTGGTCTAGGTGGCGTTGGCGGTCTAGG
TGGAGCTGGTTTAGGCGGTGTAGGTGGTGTTGGCGGTGGTATTGGTAAAGCCGGT
GGTATTGGCGGTTTAGGTGGTCTAGGCGGAGCCGGAGGTGGTTTAGGTGGAGTT
GGTGGTCTCGGTAAGGCTGGTGGTATTGGTGTTGGTGGTGGTATCGGTGGTGGAC
ACGGCGTGGTCGGTGGTGTGATCGATCCACATCCTTAA

SEQ ID NO:44, Deduced amino acid sequence of the open reading frame of Jb007
MSRALSVVCVLLAISFVCARARQVPGESDEGKTTGHDDTTTMPMHAKAADQLPPKS
VGDKKCIGGVAGVGGFAGVGGVAGVGGLGMPLIGGLGGIGKYGGIGGAAGIGGFHS
IGGVGGLGGVGGGVGGLGGVGGGVGGLGGVGGLGGAGLGGVGGVGGGIGKAGGI
GGLGGLGGAGGGLGGVGGLGKAGGIGVGGGIGGGHGVVGGVIDPHP

SEQ ID NO:45, Nucleotide sequence of the open reading frame of Jb009
ATGGCAAGCAGCGACGTGAAGCTGATCGGTGCATGGGCGAGTCCCTTTGTGATG
AGGCCGAGGATTGCTCTAAACCTCAAGTCTGTCCCCTACGAGTTCCTCCAAGAGA
CGTTTGGGTCTAAGAGCGAGTTGCTTCTTAAATCAAACCCGGTTCACAAGAAGAT
CCCGGTTCTGCTTCATGCTGATAAACCGGTGAGTGAGTCCAACATCATCGTTGAG
TATATCGATGACACTTGGAGCTCATCTGGACCGTCCATTCTCCCTTCCGATCCTTA
CGATCGGGCCATGGCTCGGTTCTGGGCTGCTTACATCGACGAAAAGTGGTTTGTC
GCTCTAAGAGGTTTCCTAAAAGCCGGAGGAGAAGAAGAGAAGAAAGCTGTGATA
GCTCAACTAGAAGAAGGGAATGCGTTTCTGGAGAAGGCGTTCATTGATTGCAGC
AAAGGAAAACCGTTCTTCAACGGTGACAACATCGGTTACCTCGACATTGCTCTCG
GGTGCTTCTTGGCTTGGTTGAGAGTCACCGAGTTAGCAGTCAGCTATAAAATTCT
TGATGAGGCCAAGACACCTTCTTTGTCCAAATGGGCTGAGAATTTCTGTAATGAT
CCCGCTGTAAAACCTGTCATGCCCGAGACTGCAAAGCTTGCTGAATTCGCAAAGA
AGATCTTTCCTAAGCCGCAGGCCTAA

SEQ ID NO:46, Deduced amino acid sequence of the open reading frame of Jb009
MASSDVKLIGAWASPFVMRPRIALNLKSVPYEFLQETFGSKSELLLKSNPVHKKIPVL
LHADKPVSESNIIVEYIDDTWSSSGPSILPSDPYDRAMARFWAAYIDEKWFVALRGFL
KAGGEEEKKAVIAQLEEGNAFLEKAFIDCSKGKPFFNGDNIGYLDIALGCFLAWLRV
TELAVSYKILDEAKTPSLSKWAENFCNDPAVKPVMPETAKLAEFAKKIFPKPQA

SEQ ID NO:47, Nucleotide sequence of the open reading frame of Jb013
ATGGCGTCTCAACAAGAGAAGAAGCAGCTGGATGAGAGGGCAAAGAAGGGCGA
GACCGTCGTGCCAGGTGGTACGGGAGGCAAAAGCTTCGAAGCTCAACAGCATCT
CGCTGAAGGGAGGAGCCGAGGAGGGCAAACTCGAAAGGAGCAGTTAGGAACTG
AAGGATATCAGCAGATGGGACGCAAAGGTGGTCTTAGCACCGGAGACAAGCCTG
GTGGGGAACACGCTGAGGAGGAAGGAGTCGAGATAGACGAATCCAAATTCAGG
ACCAAGACCTAA

SEQ ID NO:48, Deduced amino acid sequence of the open reading frame of Jb013
MASQQEKKQLDERAKKGETVVPGGTGGKSFEAQQHLAEGRSRGGQTRKEQLGTEG
YQQMGRKGGLSTGDKPGGEHAEEEGVEIDESKFRTKT

SEQ ID NO:51, Nucleotide sequence of the open reading frame of Jb017

ATGGCTCCTTCAACAAAAGTTCTCTCTTTACTTCTCTTATATGGCGTCGTGTCATT
AGCCTCCGGTGATGAGTCCATCATCAACGACCATCTCCAACTTCCATCGGACGGC
AAGTGGAGAACCGATGAAGAAGTGAGGTCCATCTACTTACAATGGTCCGCAGAA
CACGGGAAAACTAACAACAACAACAACGGTATCATCAACGACCAAGACAAAAG
ATTCAATATTTTCAAAGACAACTTAAGATTCATCGATCTACACAACGAAAACAAC
AAGAACGCTACTTACAAGCTTGGTCTCACCAAATTTACCGATCTCACTAACGATG
AGTACCGCAAGTTGTACCTCGGGGCAAGAACTGAGCCCGCCCGCCGCATCGCTA
AGGCCAAGAATGTCAACCAGAAATACTCAGCCGCTGTAAACGGCAAGGAGGTTC
CAGAGACGGTTGATTGGAGACAGAAAGGAGCCGTTAACCCCATCAAAGACCAAG
GAACTTGCGGAAGTTGTTGGGCGTTTTCGACTACTGCAGCAGTAGAAGGTATAAA
CAAGATCGTAACAGGAGAACTCATATCTCTATCAGAACAAGAACTTGTTGACTGC
GACAAATCCTACAATCAAGGTTGCAACGGCGGTTTAATGGACTACGCTTTTCAAT
TCATCATGAAAAATGGTGGCTTAAACACTGAGAAAGATTATCCTTACCGTGGATT
CGGCGGAAAATGCAATTCTTTCTTGAAGAATTCTAGAGTTGTGAGTATTGATGGG
TACGAAGATGTTCCTACTAAAGACGAGACTGCGTTGAAGAAAGCTATTTCATACC
AACCGGTTAGTGTAGCTATTGAAGCCGGTGGAAGAATTTTTCAACATTACCAATC
GGGTATTTTTACCGGAAGTTGTGGTACAAATCTTGATCACGCGGTAGTTGCTGTC
GGGTACGGATCAGAGAACGGTGTTGACTACTGGATTGTAAGGAACTCTTGGGGT
CCACGTTGGGGTGAGGAAGGTTACATTAGAATGGAGAGAAACTTGGCAGCCTCC
AAATCCGGTAAGTGTGGGATTGCGGTTGAAGCCTCGTACCCGGTTAAGTACAGCC
CAAACCCGGTTCGTGGAAATACTATCAGCAGTGTTTGA

SEQ ID NO:52, Amino acid sequence of the open reading frame of Jb017

MAPSTKVLSLLLLYGVVSLASGDESIINDHLQLPSDGKWRTDEEVRSIYLQWSAEHG
KTNNNNNGIINDQDKRFNIFKDNLRFIDLHNENNKNATYKLGLTKFTDLTNDEYRKL
YLGARTEPARRIAKAKNVNQKYSAAVNGKEVPETVDWRQKGAVNPIKDQGTCGSC
WAFSTTAAVEGINKIVTGELISLSEQELVDCDKSYNQGCNGGLMDYAFQFIMKNGGL
NTEKDYPYRGFGGKCNSFLKNSRVVSIDGYEDVPTKDETALKKAISYQPVSVAIEAG
GRIFQHYQSGIFTGSCGTNLDHAVVAVGYGSENGVDYWIVRNSWGPRWGEEGYIRM
ERNLAASKSGKCGIAVEASYPVKYSPNPVRGNTISSV

SEQ ID NO:53, Nucleotide sequence of the open reading frame of Jb024

ATGCGGTGCTTTCCACCTCCCTTATGGTGCACCTCCTTGGTCGTTTTCTTGTCGGT
TACCGGAGCCCTAGCCGCCGATCCCTACGTCTTCTTCGATTGGACTGTCTCTTACC
TCTCTGCTTCTCCTCTCGGCACTCGTCAACAGGTAATTGGGATAAATGGGCAATT
TCCTGGTCCGATTCTAAACGTAACTACGAATTGGAATGTTGTTATGAATGTGAAG
AATAATCTTGATGAGCCATTGCTTCTTACATGGAATGGAATCCAACATAGGAAAA
ACTCATGGCAAGATGGTGTTTTGGGAACTAATTGTCCAATTCCTTCTGGTTGGAA
TTGGACTTATGAGTTTCAAGTTAAAGATCAGATTGGTAGTTTCTTTTATTTTCCTT
CTACAAATTTTCAAAGAGCTTCTGGTGGTTATGGAGGGATTATTGTCAATAATCG
CGCTATCATTCCGGTTCCTTTCGCTCTTCCTGATGGTGATGTTACTCTCTTTATCAG
TGATTGGTATACTAAGAGCCATAAGAAGCTGAGGAAGGATGTTGAGAGTAAGAA
CGGCCTTCGACCTCCGGATGGTATTGTCATCAATGGATTTGGACCTTTTGCTTCTA
ATGGTAGTCCTTTTGGGACCATAAACGTTGAACCAGGACGAACATATCGTTTTCG
TGTTCACAATAGTGGCATTGCGACCAGCTTGAATTTCAGAATACAGAATCATAAC
CTGCTTCTTGTTGAGACAGAAGGGTCATACACAATTCAGCAGAATTATACGAATA
TGGATATACATGTGGGTCAATCTTTCTCATTTCTGGTCACTATGGATCAGTCTGGT
AGTAATGACTACTACATTGTTGCCAGCCCAAGGTTTGCTACATCCATCAAAGCTA
GTGGAGTCGCTGTCTTGCGCTACTCTAATTCCCAAGGACCCGCTTCAGGTCCACT

CCCTGATCCTCCTATTGAGTTGGACACATTTTTCTCAATGAACCAAGCACGATCCT
TAAGGTTGAATTTGTCATCTGGAGCTGCCCGTCCAAACCCGCAGGGATCTTTCAA
ATATGGCCAGATTACAGTAACTGATGTGTATGTGATTGTCAACCGACCACCAGAG
ATGATAGAGGGACGATTGCGTGCAACTCTTAATGGTATATCATACTTACCTCCTG
CAACACCCCTAAAGCTTGCTCAGCAATACAACATCTCAGGGGTATACAAGTTGG
ATTTCCCAAAAAGGCCAATGAATAGGCACCCCAGGGTTGATACCTCAGTCATAA
ACGGCACGTTCAAGGGATTCGTGGAAATCATATTTCAAAATAGTGACACCACTGT
TAAGAGCTACCACTTGGATGGTTATGCATTTTTTGTTGTTGGGATGGACTTTGGTC
TGTGGACAGAAAATAGCAGAAGCACATACAACAAGGGTGATGCAGTTGCTCGAT
CTACTACGCAGGTGTTTCCTGGTGCATGGACGGCCGTCTTGGTTTCTTTGGACAAT
GCTGGCATGTGGAACCTTCGAATAGACAATCTAGCCTCATGGTATCTTGGCCAAG
AACTATACTTGAGTGTGGTTAATCCAGAGATTGACATTGACTCATCTGAGAATTC
CGTTCCTAAAAACTCTATATATTGTGGTCGGCTCTCACCATTACAAAAGTAA

SEQ ID NO:54, Deduced amino acid sequence of the open reading frame of Jb024
MRCFPPPLWCTSLVVFLSVTGALAADPYVFFDWTVSYLSASPLGTRQQVIGINGQFP
GPILNVTTNWNVVMNVKNNLDEPLLLTWNGIQHRKNSWQDGVLGTNCPIPSGWNW
TYEFQVKDQIGSFFYFPSTNFQRASGGYGGIIVNNRAIIPVPFALPDGDVTLFISDWYT
KSHKKLRKDVESKNGLRPPDGIVINGFGPFASNGSPFGTINVEPGRTYRFRVHNSGIA
TSLNFRIQNHNLLLVETEGSYTIQQNYTNMDIHVGQSFSFLVTMDQSGSNDYYIVASP
RFATSIKASGVAVLRYSNSQGPASGPLPDPPIELDTFFSMNQARSLRLNLSSGAARPNP
QGSFKYGQITVTDVYVIVNRPPEMIEGRLRATLNGISYLPPATPLKLAQQYNISGVYK
LDFPKRPMNRHPRVDTSVINGTFKGFVEIIFQNSDTTVKSYHLDGYAFFVVGMDFGL
WTENSRSTYNKGDAVARSTTQVFPGAWTAVLVSLDNAGMWNLRIDNLASWYLGQ
ELYLSVVNPEIDIDSSENSVPKNSIYCGRLSPLQK

SEQ ID NO:55, Nucleotide sequence of the open reading frame of Jb027
ATGCTTCTAATTCTAGCGATTTGGTCACCAATTTCACACTCGCTTCACTTCGATCT
ACACTCAGGTCGCACAAAGTGTATCGCCGAAGACATCAAAAGCAATTCAATGAC
TGTTGGTAAATACAACATCGATAATCCTCACGAAGGTCAAGCTTTACCACAAACT
CACAAAATTTCCGTCAAGGTGACGTCTAATTCCGGTAACAATTACCATCACGCGG
AACAAGTAGATTCAGGACAATTCGCATTCTCGGCTGTTGAAGCAGGTGATTACAT
GGCTTGTTTCACTGCTGTTGATCATAAGCCTGAGGTTTCGTTGAGTATTGACTTTG
AGTGGAAGACTGGTGTTCAATCTAAAAGCTGGGCTAATGTTGCTAAGAAGAGTC
AAGTCGAAGTTATGGAATTTGAAGTAAAGAGTCTTCTTGATACTGTTAACTCGAT
TCATGAAGAGATGTATTATCTTAGAGATAGGGAAGAAGAGATGCAAGACTTGAA
CCGGTCCACTAACACAAAAATGGCGTGGTTGAGTGTTCTCTCGTTTTTCGTCTGC
ATAGGAGTTGCAGGGATGCAGTTTTTGCACTTGAAGACGTTTTTCGAGAAGAAGA
AGGTTATCTGA

SEQ ID NO:56, Deduced amino acid sequence of the open reading frame of Jb027
MLLILAIWSPISHSLHFDLHSGRTKCIAEDIKSNSMTVGKYNIDNPHEGQALPQTHKIS
VKVTSNSGNNYHHAEQVDSGQFAFSAVEAGDYMACFTAVDHKPEVSLSIDFEWKT
GVQSKSWANVAKKSQVEVMEFEVKSLLDTVNSIHEEMYYLRDREEEMQDLNRSTN
TKMAWLSVLSFFVCIGVAGMQFLHLKTFFEKKKVI

SEQ ID NO:57, Nucleotide sequence of the open reading frame of OO-1
ATGGCACATGCCACGTTTACGTCGGAAGGGCAGAATATGGAGTCGTTTCGACTCT
TGAGTGGCCACAAAATCCCAGCCGTTGGACTCGGCACGTGGCGATCTGGGTCTCA
AGCCGCCCACGCCGTTGTCACTGCAATCGTCGAGGGTGGCTATAGGCACATAGAT

ACAGCTTGGGAGTATGGTGATCAGAGAGAGGTCGGTCAAGGAATAAAGAGGGC
GATGCACGCTGGCCTTGAAAGGAGGGACCTCTTTGTGACCTCGAAGCTTTGGTGC
ACTGAGTTATCTCCTGAGAGAGTGCGTCCTGCTCTGCAAAACACCCTTAAAGAGC
TTCAATTAGAGTACCTTGATCTCTACTTGATTCACTGGCCTATCCGGCTAAGAGA
AGGAGCCAGTAAGCCACCAAAGGCAGGGGACGTTCTTGACTTTGACATGGAAGG
AGTTTGGAGAGAAATGGAGAATCTTTCCAAGGACAGTCTCGTCAGGAATATCGG
TGTCTGTAACTTTACAGTCACTAAGCTCAATAAGCTGCTAGGATTTGCTGAACTG
ATCCCTGCCGTTTGCCAGATGGAAATGCATCCTGGTTGGAGAAACGATAGGATAC
TCGAATTCTGCAAGAAGAATGAGATCCATGTTACTGCCTATTCTCCATTGGGATC
TCAAGAAGGCGGGAGAGATCTGATACACGATCAGACGGTGGATAGGATAGCGA
AGAAGCTGAATAAGACACCGGGACAGATTCTAGTGAAATGGGGTTTGCAGAGAG
GAACAAGTGTCATCCCTAAGTCATTGAATCCAGAGAGGATCAAAGAGAACATCA
AAGTGTTTGATTGGGTGATCCCTGAACAAGACTTCCAAGCTCTCAACAGCATCAC
TGACCAGAAACGAGTGATAGACGGTGAGGATCTTTTCGTCAACAAGACCGAAGG
TCCATTCCGTAGTGTGGCTGATCTATGGGACCATGAAGACTAA

SEQ ID NO:58, Deduced amino acid sequence of the open reading frame of OO-1
MAHATFTSEGQNMESFRLLSGHKIPAVGLGTWRSGSQAAHAVVTAIVEGGYRHIDT
AWEYGDQREVGQGIKRAMHAGLERRDLFVTSKLWCTELSPERVRPALQNTLKELQL
EYLDLYLIHWPIRLREGASKPPKAGDVLDFDMEGVWREMENLSKDSLVRNIGVCNF
TVTKLNKLLGFAELIPAVCQMEMHPGWRNDRILEFCKKNEIHVTAYSPLGSQEGGRD
LIHDQTVDRIAKKLNKTPGQILVKWGLQRGTSVIPKSLNPERIKENIKVFDWVIPEQDF
QALNSITDQKRVIDGEDLFVNKTEGPFRSVADLWDHED

SEQ ID NO:59, Nucleotide sequence of the open reading frame of OO-2
ATGGCGTCTGAGAAACAAAAACAACATGCACAACCTGGCAAAGAACATGTCATG
GAATCAAGCCCACAATTCTCTAGCTCAGATTACCAACCTTCCAACAAGCTTCGTG
GTAAGGTGGCGTTGATAACTGGTGGAGACTCTGGGATTGGTCGAGCCGTGGGAT
ACTGTTTTGCATCCGAAGGAGCTACTGTGGCTTTCACTTACGTGAAGGGTCAAGA
AGAAAAAGATGCACAAGAGACCCTACAAATGTTGAAGGAGGTCAAAACCTCGG
ACTCCAAGGAACCTATCGCCATTCCAACGGATTTAGGATTTGACGAAAACTGCAA
AAGGGTCGTTGATGAGGTCGTTAATGCTTTTGGCCGCATCGATGTTTTGATCAAT
AACGCAGCAGAGCAGTACGAGAGCAGCACAATCGAAGAGATTGATGAGCCTAG
GCTTGAGCGAGTCTTCCGTACAAACATCTTTTCTTACTTCTTTCTCACAAGGCATG
CGTTGAAGCATATGAAGGAAGGAAGCAGCATTATCAACACCACTTCGGTGAATG
CCTACAAGGGAAACGCTTCACTTCTCGACTACACCGCTACAAAAGGAGCGATTGT
GGCGTTTACTCGAGGACTTGCACTTCAGCTAGCTGAGAAAGGAATCCGTGTCAAT
GGTGTGGCTCCTGGTCCAATATGGACACCCCTTATCCCAGCATCATTCAATGAGG
AGAAGATTAAGAATTTTGGGTCTGAGGTTCCGATGAAAAGAGCGGGTCAGCCAA
TTGAAGTGGCACCATCCTATGTTTTCTTGGCGTGTAACCACTGCTCTTCTTACTTC
ACTGGTCAAGTTCTTCACCCTAATGGAGGAGCTGTGGTAAATGCGTAA

SEQ ID NO:60, Deduced amino acid sequence of the open reading frame of OO-2
MASEKQKQHAQPGKEHVMESSPQFSSSDYQPSNKLRGKVALITGGDSGIGRAVGYC
FASEGATVAFTYVKGQEEKDAQETLQMLKEVKTSDSKEPIAIPTDLGFDENCKRVVD
EVVNAFGRIDVLINNAAEQYESSTIEEIDEPRLERVFRTNIFSYFFLTRHALKHMKEGS
SIINTTSVNAYKGNASLLDYTATKGAIVAFTRGLALQLAEKGIRVNGVAPGPIWTPLIP
ASFNEEKIKNFGSEVPMKRAGQPIEVAPSYVFLACNHCSSYFTGQVLHPNGGAVVNA

SEQ ID NO:61, Nucleotide sequence of the open reading frame of OO-3

ATGGATTCAACGAAGCTTAGTGAGCTAAAGGTCTTCATCGATCAATGCAAGTCTG
ACCCTTCCCTTCTCACTACTCCTTCACTCTCCTTCTTCCGTGACTATCTCGAGAGTC
TTGGTGCTAAGATACCTACTGGTGTCCATGAAGAAGACAAAGACACTAAGCCGA
GGAGTTTCGTAGTGGAAGAGAGTGATGATGATATGGATGAAACTGAAGAAGTAA
AACCGAAAGTGGAGGAAGAAGAAGAAGAGGATGAGATTGTTGAATCTGATGTA
GAGCTTGAAGGAGACACTGTTGAGCCTGATAATGATCCTCCTCAGAAGATGGGG
GATTCATCAGTGGAGGTGACTGATGAGAATCGTGAAGCTGCTCAAGAAGCTAAG
GGCAAAGCCATGGAGGCCCTTTCTGAAGGAAACTTTGATGAAGCAATTGAGCAT
TTAACTCGGGCAATAACGTTGAACCCGACTTCAGCTATTATGTATGGAAACAGAG
CTAGTGTCTACATTAAGTTGAAGAAGCCAAACGCTGCTATTCGAGATGCAAACGC
AGCATTGGAGATTAACCCTGATTCTGCCAAGGGATACAAGTCACGAGGTATGGC
TCGTGCCATGCTTGGAGAATGGGCAGAGGCTGCAAAAGACCTTCACCTTGCATCT
ACGATAGACTATGATGAGGAAATTAGTGCTGTTCTCAAAAAGGTTGAACCTAAT
GCACATAAGCTTGAGGAGCACCGTAGAAAGTATGACAGATTACGTAAGGAAAGA
GAGGACAAAAAGGCTGAACGGGATAGATTACGTCGCCGTGCTGAAGCACAGGCT
GCCTATGATAAAGCTAAGAAAGAAGAACAGTCATCATCTAGCAGACCATCAGGA
GGCGGTTTCCCAGGAGGTATGCCCGGTGGTTTCCCAGGAGGTATGCCCGGTGGAT
TCCCAGGAGGAATGGGAGGCATGCCCGGCGGATTCCCGGGAGGAATGGGTGGTA
TGGGCGGTATGCCCGGTGGATTCCCAGGAGGAATGGGCGGTGGTATGCCTGCAG
GAATGGGCGGTGGTATGCCCGGAATGGGCGGTGGTATGCCTGCTGGAATGGGTG
GTGGCGGTATGCCAGGTGCAGGCGGTGGTATGCCTGGTGGTGGCGGTATGCCTG
GTGGTATGGACTTCAGCAAAATATTGAATGATCCTGAGCTAATGACGGCATTTAG
CGACCCTGAAGTCATGGCTGCTCTTCAAGATGTGATGAAGAACCCTGCGAATCTA
GCGAAGCATCAGGCGAATCCGAAGGTGGCTCCCGTGATTGCAAAGATGATGGGC
AAATTTGCAGGACCTCAGTAA

SEQ ID NO:62, Deduced amino acid sequence of the open reading frame of OO-3

MDSTKLSELKVFIDQCKSDPSLLTTPSLSFFRDYLESLGAKIPTGVHEEDKDTKPRSFV
VEESDDDMDETEEVKPKVEEEEEEDEIVESDVELEGDTVEPDNDPPQKMGDSSVEVT
DENREAAQEAKGKAMEALSEGNFDEAIEHLTRAITLNPTSAIMYGNRASVYIKLKKP
NAAIRDANAALEINPDSAKGYKSRGMARAMLGEWAEAAKDLHLASTIDYDEEISAV
LKKVEPNAHKLEEHRRKYDRLRKEREDKKAERDRLRRRAEAQAAYDKAKKEEQSS
SSRPSGGGFPGGMPGGFPGGMPGGFPGGMGGMPGGFPGGMGGMGGMPGGFPGGM
GGGMPAGMGGGMPGMGGGMPAGMGGGGMPGAGGGMPGGGGMPGGMDFSKILN
DPELMTAFSDPEVMAALQDVMKNPANLAKHQANPKVAPVIAKMMGKFAGPQ

SEQ ID NO:63, Nucleotide sequence of the open reading frame of OO-4

ATGAAGGTTCACGAGACAAGATCTCACGCTCACATGTCTGGAGACGAACAAAAG
AAGGGAAATTTGCGGAAGCACAAAGCAGAAGGGAAACTTCCAGAATCTGAACA
GTCTCAGAAGAAGGCAAAGCCTGAAAACGATGACGGACGTTCTGTCAACGGCGC
CGGAGATGCTGCTTCAGAGTACAATGAGTTCTGCAAAGCGGTTGAGGAGAATCT
GTCCATTGATCAGATTAAAGAAGTTCTCGAAATCAACGGCCAAGATTGTTCTGCT
CCAGAAGAGACCTTGCTAGCTCAATGTCAAGATTTGCTGTTCTATGGGGCATTAG
CTAAATGTCCTTTATGCGGAGGAACTTTAATTTGCGACAATGAAAAGAGATTTGT
ATGTGGAGGTGAGATAAGTGAGTGGTGCAGTTGCGTGTTTAGTACGAAAGATCC
TCCTAGAAAGGAAGAGCCAGTTAAAATCCCTGATTCTGTCATGAACTCTGCTATA
TCTGACTTGATCAAGAAACACCAGGACCCTAAAAGCCGACCTAAAAGAGAGTTA
GGCTCTGCTGATAAACCCTTTGTGGGAATGATGATCTCTCTCATGGGACGTCTCA
CGAGAACACATCAATATTGGAAGAAAAAGATCGAGAGAAACGGTGGGAAAGTC

TCCAATACTGTTCAAGGCGTAACATGTTTGGTGGTTTCGCCAGCTGAAAGAGAAC
GAGGTGGTACGTCAAAGATGGTGGAGGCAATGGAACAAGGTCTACCGGTTGTGA
GCGAAGCATGGTTGATCGACAGCGTGGAGAAGCATGAAGCTCAGCCACTTGAAG
CTTATGACGTGGTCAGTGATCTTTCAGTGGAAGGGAAAGGAATTCCATGGGATA
AGCAAGATCCTAGTGAGGAGGCAATTGAATCCTTTTCTGCTGAGCTCAAAATGTA
TGGGAAAAGAGGAGTGTACATGGACACAAAACTTCAGGAGAGAGGAGGAAAGA
TCTTCGAGAAAGATGGACTCTTGTATAACTGTGCCTTCTCGATATGCGATTTGGG
AAAAGGGCGTAATGAGTATTGTATTATGCAGCTAGTCACGGTACCCGATAGTAA
CCTGAACATGTACTTCAAGAGAGGGAAAGTAGGAGATGACCCTAATGCCGAAGA
GAGGCTCGAGGAATGGGAGGACGAAGAAGCTGCGATCAAAGAGTTTGCAAGGC
TTTTTGAGGAGATAGCAGGGAATGAGTTTGAGCCATGGGAACGTGAGAAGAAGA
TTCAAAAGAAGCCTCATAAGTTTTTCCCAATTGATATGGATGATGGAATCGAAGT
AAGGAGTGGGGCTCTTGGTCTAAGGCAGCTTGGCATTGCTTCTGCTCATTGCAAG
CTTGATTCGTTTGTTGCAAACTTCATTAAAGTTCTGTGTGGTCAAGAGATTTACAA
TTACGCGTTGATGGAGCTTGGATTGGATCCGCCCGATCTACCTATGGGAATGCTA
ACTGATATCCACTTGAAACGATGCGAAGAGGTATTACTCGAGTTTGTTGAGAAGG
TCAAAACAACAAAAGAGACAGGTCAGAAAGCTGAAGCAATGTGGGCAGACTTC
AGCTCACGATGGTTCTCTTTGATGCACAGCACTAGGCCGATGCGATTACACGATG
TCAATGAACTTGCAGACCATGCGGCCTCTGCTTTTGAGACGGTGAGGGACATAAA
CACAGCATCTCGTTTGATAGGGGACATGCGAGGAGACACACTCGATGATCCGTT
GTCTGATAGGTACAAAAAACTTGGCTGCAAGATATCTGTGGTAGACAAAGAGTC
TGAAGATTACAAGATGGTTGTGAAGTATCTCGAGACTACTTATGAGCCTGTGAAA
GTCTCTGATGTTGAGTACGGTGTGTCAGTGCAGAATGTTTTTGCGGTTGAGTCAG
ATGCAATTCCTTCATTAGATGATATCAAGAAGTTACCAAATAAGGTCCTTTTATG
GTGTGGGTCTCGGAGCTCAAATCTATTGAGACATATCTACAAAGGGTTCTTACCT
GCTGTATGCTCTCTTCCGGTTCCTGGTTATATGTTTGGGAGAGCGATAGTGTGTTC
AGATGCAGCTGCAGAAGCAGCAAGGTATGGTTTTACGGCTGTGGATAGACCAGA
AGGGTTTCTTGTATTAGCCGTAGCATCACTTGGTGAGGAAGTTACAGAATTTACA
AGTCCACCAGAGGATACGAAGACGTTGGAAGATAAAAAGATTGGAGTGAAAGG
ATTAGGGAGGAAGAAAACTGAAGAGTCGGAGCATTTCATGTGGAGAGATGACAT
AAAAGTTCCTTGTGGACGGTTGGTTCCATCGGAACATAAGGACAGTCCACTTGAG
TACAACGAGTACGCGGTTTATGATCCGAAACAGACAAGTATAAGGTTCTTGGTG
GAAGTGAAGTACGAGGAGAAGGGAACTGAGATAGTCGATGTCGAACCAGAGTA
G

SEQ ID NO:64, Deduced amino acid sequence of the open reading frame of OO-4

MKVHETRSHAHMSGDEQKKGNLRKHKAEGKLPESEQSQKKAKPENDDGRSVNGA
GDAASEYNEFCKAVEENLSIDQIKEVLEINGQDCSAPEETLLAQCQDLLFYGALAKCP
LCGGTLICDNEKRFVCGGEISEWCSCVFSTKDPPRKEEPVKIPDSVMNSAISDLIKKHQ
DPKSRPKRELGSADKPFVGMMISLMGRLTRTHQYWKKKIERNGGKVSNTVQGVTCL
VVSPAERERGGTSKMVEAMEQGLPVVSEAWLIDSVEKHEAQPLEAYDVVSDLSVEG
KGIPWDKQDPSEEAIESFSAELKMYGKRGVYMDTKLQERGGKIFEKDGLLYNCAFSI
CDLGKGRNEYCIMQLVTVPDSNLNMYFKRGKVGDDPNAEERLEEWEDEEAAIKEFA
RLFEEIAGNEFEPWEREKKIQKKPHKFFPIDMDDGIEVRSGALGLRQLGIASAHCKLD
SFVANFIKVLCGQEIYNYALMELGLDPPDLPMGMLTDIHLKRCEEVLLEFVEKVKTT
KETGQKAEAMWADFSSRWFSLMHSTRPMRLHDVNELADHAASAFETVRDINTASR
LIGDMRGDTLDDPLSDRYKKLGCKISVVDKESEDYKMVVKYLETTYEPVKVSDVEY
GVSVQNVFAVESDAIPSLDDIKKLPNKVLLWCGSRSSNLLRHIYKGFLPAVCSLPVPG
YMFGRAIVCSDAAAEAARYGFTAVDRPEGFLVLAVASLGEEVTEFTSPPEDTKTLED

KKIGVKGLGRKKTEESEHFMWRDDIKVPCGRLVPSEHKDSPLEYNEYAVYDPKQTSI
RFLVEVKYEEKGTEIVDVEPE

SEQ ID NO:65, Nucleotide sequence of the open reading frame of OO-5

ATGTCTACCCCAGCTGAATCTTCAGACTCGAAATCGAAGAAAGATTTCAGTACTG
CTATTCTCGAGAGGAAGAAGTCTCCGAACCGTCTCGTCGTCGATGAGGCTATCAA
CGATGATAACTCCGTCGTCTCTCTTCACCCTGCAACCATGGAGAAGCTTCAGCTC
TTCCGTGGTGATACCATTCTCATCAAGGGTAAGAAGAGGAAGGACACTGTCTGC
ATTGCTCTTGCTGATGAGACATGTGAGGAGCCAAAGATCAGAATGAATAAAGTA
GTCAGATCTAACTTGAGGGTTAGACTGGGAGATGTTATATCTGTTCACCAATGCC
CAGACGTCAAGTACGGAAAGCGTGTTCACATCCTGCCTGTTGATGATACTGTTGA
AGGAGTGACTGGAAACCTATTTGATGCTTACCTGAAACCTTATTTCCTTGAGGCA
TACCGTCCAGTGAGGAAGGGTGATCTCTTCCTAGTCAGAGGAGGAATGAGGAGT
GTGGAGTTCAAAGTTATAGAGACAGATCCTGCTGAGTACTGCGTGGTTGCTCCAG
ACACAGAGATTTTCTGTGAGGGTGAGCCTGTGAAGAGAGAGGATGAAGAAAGGC
TAGATGATGTAGGTTATGATGATGTTGGTGGTGTCAGGAAACAGATGGCTCAGAT
TAGGGAACTTGTTGAACTTCCCTTGAGGCATCCACAGCTATTCAAGTCGATTGGT
GTTAAGCCACCGAAGGGAATTCTTCTTTATGGACCACCTGGGTCTGGAAAGACTT
TGATCGCTCGTGCTGTGGCTAATGAAACGGGTGCCTTTTTCTTCTGTATCAACGG
ACCTGAGATCATGTCCAAATTGGCTGGTGAGAGTGAGAGCAACCTCAGGAAAGC
ATTCGAGGAGGCTGAGAAAAATGCGCCTTCAATCATATTCATTGATGAGATCGAC
TCTATTGCACCGAAAAGAGAGAAGACTAATGGAGAGGTTGAGAGGAGGATTGTC
TCTCAGCTCCTTACGCTAATGGATGGACTGAAATCTCGTGCTCATGTTATCGTCAT
GGGAGCAACCAATCGCCCCAACAGTATCGACCCAGCTTTGAGAAGGTTTGGAAG
ATTTGACAGGGAGATCGATATTGGAGTTCCTGACGAAATTGGACGTCTTGAAGTT
CTGAGGATCCATACAAAGAACATGAAGCTGGCTGAAGATGTGGATCTCGAAAGG
ATCTCAAAGGACACACACGGTTACGTCGGTGCTGATCTTGCAGCTTTGTGCACAG
AGGCCGCCCTGCAATGCATCAGGGAGAAGATGGATGTGATTGATCTGGAAGATG
ACTCCATAGACGCTGAAATCCTCAATTCCATGGCAGTCACTAATGAACATTTCCA
CACTGCTCTCGGGAACAGCAACCCATCTGCACTTCGTGAAACTGTTGTGGAGGTT
CCCAACGTCTCTTGGAATGATATTGGAGGTCTTGAGAATGTCAAGAGAGAGCTCC
AGGAGACTGTTCAATACCCAGTCGAGCACCCAGAGAAGTTTGAGAAATTCGGGA
TGTCTCCATCAAAGGGAGTCCTTTTCTACGGTCCTCCTGGATGTGGGAAAACCCT
TTTGGCCAAAGCTATTGCCAACGAGTGCCAAGCTAATTTCATCAGTGTCAAGGGT
CCCGAGCTTCTGACAATGTGGTTTGGAGAGAGTGAAGCAAATGTTCGTGAAATCT
TCGACAAGGCCCGTCAATCCGCTCCATGTGTTCTTTTCTTTGATGAGCTCGACTCC
ATTGCAACTCAGAGAGGAGGTGGAAGTGGTGGCGATGGAGGTGGTGCTGCGGAC
AGAGTCTTGAACCAGCTTTTGACTGAGATGGACGGAATGAATGCCAAGAAAACC
GTCTTCATCATCGGAGCTACCAACAGACCTGACATTATCGATTCAGCTCTTCTCC
GTCCTGGAAGGCTTGACCAGCTCATTTACATTCCACTACCAGATGAGGATTCCCG
TCTCAATATCTTCAAGGCCGCCTTGAGGAAATCTCCTATTGCTAAAGATGTAGAC
ATCGGTGCACTTGCTAAATACACTCAGGGTTTCAGTGGTGCTGATATCACTGAGA
TTTGCCAGAGAGCTTGCAAGTACGCCATCAGAGAAAACATTGAGAAGGACATTG
AAAAGGAGAAGAGGAGGAGCGAGAACCCAGAGGCAATGGAGGAAGATGGAGT
GGATGAAGTATCAGAGATCAAAGCTGCACACTTTGAGGAGTCGATGAAGTATGC
GCGTAGGAGTGTGAGTGATGCAGACATCAGGAAGTACCAAGCCTTTGCTCAGAC
GTTGCAGCAGTCTAGAGGGTTCGGTTCTGAGTTCAGGTTCGAGAATTCTGCTGGT
TCAGGTGCCACCACTGGAGTCGCAGATCCGTTTGCCACGTCTGCAGCCGCTGCTG
GGGACGATGATGATCTCTACAATTAG

SEQ ID NO:66, Deduced amino acid sequence of the open reading frame of OO-5

MSTPAESSDSKSKKDFSTAILERKKSPNRLVVDEAINDDNSVVSLHPATMEKLQLFRG
DTILIKGKKRKDTVCIALADETCEEPKIRMNKVVRSNLRVRLGDVISVHQCPDVKYG
KRVHILPVDDTVEGVTGNLFDAYLKPYFLEAYRPVRKGDLFLVRGGMRSVEFKVIET
DPAEYCVVAPDTEIFCEGEPVKREDEERLDDVGYDDVGGVRKQMAQIRELVELPLR
HPQLFKSIGVKPPKGILLYGPPGSGKTLIARAVANETGAFFFCINGPEIMSKLAGESES
NLRKAFEEAEKNAPSIIFIDEIDSIAPKREKTNGEVERRIVSQLLTLMDGLKSRAHVIV
MGATNRPNSIDPALRRFGRFDREIDIGVPDEIGRLEVLRIHTKNMKLAEDVDLERISK
DTHGYVGADLAALCTEAALQCIREKMDVIDLEDDSIDAEILNSMAVTNEHFHTALGN
SNPSALRETVVEVPNVSWNDIGGLENVKRELQETVQYPVEHPEKFEKFGMSPSKGVL
FYGPPGCGKTLLAKAIANECQANFISVKGPELLTMWFGESEANVREIFDKARQSAPC
VLFFDELDSIATQRGGGSGGDGGGAADRVLNQLLTEMDGMNAKKTVFIIGATNRPDI
IDSALLRPGRLDQLIYIPLPDEDSRLNIFKAALRKSPIAKDVDIGALAKYTQGFSGADIT
EICQRACKYAIRENIEKDIEKEKRRSENPEAMEEDGVDEVSEIKAAHFEESMKYARRS
VSDADIRKYQAFAQTLQQSRGFGSEFRFENSAGSGATTGVADPFATSAAAAGDDDD
LYN

SEQ ID NO:67, Nucleotide sequence of the open reading frame of OO-6

ATGGACAAATCTAGTACCATGCTTGTTCACTATGACAAAGGGACTCCAGCAGTTG
CTAATGAGATTAAAGAAGCTCTCGAAGGAAATGATGTTGAAGCTAAAGTTGATG
CCATGAAGAAGGCAATTATGCTTTTGCTGAATGGTGAAACCATTCCTCAGCTTTT
CATTACCATTATAAGATATGTGCTGCCTTCTGAAGACCACACCATCCAAAAGCTT
CTGTTGCTGTACCTGGAGCTGATTGAAAAGACAGATTCGAAGGGGAAGGTGTTG
CCTGAAATGATTTTGATATGCCAGAATCTTCGTAATAACCTTCAGCATCCGAATG
AGTACATCCGTGGAGTGACACTGAGGTTTCTCTGTCGGATGAAGGAGACTGAAA
TAGTGGAACCTTTGACTCCATCAGTGTTACAAAATCTGGAGCATCGCCATCCATT
TGTTCGCAGGAATGCAATTCTGGCAATCATGTCGATATATAAACTTCCACATGGC
GACCAACTCTTCGTGGATGCACCTGAAATGATCGAGAAAGTTCTATCAACAGAA
CAAGATCCTTCTGCCAAGAGAAATGCATTTCTAATGCTCTTTACCTGTGCCGAAG
AACGTGCAGTGAATTATCTTCTGAGCAATGTTGACAAGGTTTCAGACTGGAATGA
ATCACTTCAGATGGTGGTGCTGGAGCTGATTCGAAGTGTGTGTAAGACTAAACCA
GCGGAGAAGGGAAAATATATTAAAATTATTATTTCTCTGTTAAGTGCTACTTCTT
CTGCAGTTATCTATGAATGTGCTGGGACACTTGTTTCTCTCTCATCTGCCCCTACT
GCTATTCGAGCTGCTGCCAACACCTACTGCCAACTTCTTCTTTCTCAGAGTGACA
ACAATGTGAAGCTTATCTTGCTCGATCGGTTGTATGAGCTTAAGACATTGCACAG
AGATATCATGGTTGAGCTGATAATCGATGTGCTCAGAGCACTCTCAAGCCCAAAC
CTTGATATCCGCAGGAAGACACTTGACATTGCCCTTGACTTGATTACCCATCATA
ATATTAATGAAGTCGTTCAAATGTTGAAGAAAGAAGTTGTGAAGACACAGAGTG
GAGAACTTGAGAAGAATGGAGAGTACAGGCAAATGCTTATTCAAGCCATCCATG
CTTGTGCAGTTAAGTTCCCCGAAGTTGCAAGCACAGTGGTCCATCTTCTGATGGA
TTTCCTGGGAGATAGCAACGTGGCTTCAGCTCTTGACGTGGTTGTTTTCGTTAGA
GAGATAATAGAAACAAATCCCAAGTTGAGAGTTTCAATCATCACCAGGTTGTTG
GACACGTTCTATCAGATCCGTGCAGGAAAGGTCTGCCCTTGTGCACTTTGGATCA
TTGGTGAGTATTGCCTATCACTTTCAGAAGTTGAGAGTGGCATTTCAACTATTAC
ACAATGCCTTGGCGAATTACCATTTTACTCTGTTTCTGAGGAGTCTGAGCCAACT
GAGACATCAAAGAAGATTCAGCCTACCTCTTCTGCCATGGTGTCCTCTAGAAAGC
CAGTTATTCTTGCTGATGGAACTTATGCTACACAAAGCGCAGCCTCTGAAACCAC
ATTCTCCTCGCCTACAGTTGTTCAAGGATCACTGACTTCTGGAAATTTGAGGGCA
CTCCTTCTAACTGGTGATTTTTTCCTCGGAGCTGTGGTTGCTTGCACGTTGACCAA
ACTTGTTCTTAGGTTGGAAGAGGTTCAGTCTTCCAAAACTGAAGTAAACAAGACA

GTATCACAGGCTTTGCTAATCATGGTTTCTATTTTGCAACTTGGGCAATCTCCTGT
TTCTCCACACCCTATTGATAATGATTCGTATGAGCGGATTATGTTGTGCATAAAA
TTGCTTTGCCATAGGAATGTTGAGATGAAAAAGATATGGTTGGAATCCTGCCGCC
AGAGTTTTGTCAAGATGATTTCTGAAAAACAGCTTAGAGAGATGGAGGAACTGA
AGGCAAAGACCCAAACAACTCATGCTCAACCGGATGATCTAATTGACTTCTTCCA
TCTAAAGAGTCGGAAGGGAATGAGTCAACTTGAGTTGGAAGACCAGGTACAAGA
TGACCTAAAGCGTGCAACTGGAGAATTCACCAAGGACGAGAACGATGCTAACAA
ACTTAACCGCATTCTTCAACTCACAGGATTCAGTGACCCAGTCTATGCTGAAGCA
TATGTAACGGTACACCATTATGATATTGCTCTTGAAGTTACAGTAATCAACCGAA
CCAAGGAAACCCTTCAGAACTTGTGCTTGGAGTTAGCAACCATGGGTGATCTCAA
ACTTGTTGAGCGTCCTCAGAACTATAGTCTGGCACCTGAAAGAAGCATGCAGATT
AAAGCAAACATCAAGGTCTCGTCCACAGAGACAGGAGTCATATTCGGGAACATC
GTCTATGAGACATCAAATGTAATGGAGCGCAATGTTGTGGTTCTTAACGACATAC
ACATTGATATCATGGACTATATCTCCCCTGCTGTGTGCTCAGAGGTTGCTTTCAGA
ACTATGTGGGCAGAGTTTGAATGGGAAAACAAGGTTGCTGTGAACACCACAATT
CAAAACGAAAGAGAATTCCTCGACCACATTATCAAATCCACAAACATGAAATGT
CTCACTGCTCCATCTGCAATAGCAGGTGAATGTGGATTCCTTGCAGCAAACTTAT
ATGCAAAAAGTGTATTTGGTGAGGATGCTCTTGTGAATTTGAGTATTGAGAAGCA
AACGGATGGAACATTGAGTGGTTACATAAGGATAAGGAGCAAGACGCAAGGGA
TTGCTCTAAGTCTTGGAGACAAAATCACCCTCAAACAAAAGGGTGGTAGCTGA

SEQ ID NO:68, Deduced amino acid sequence of the open reading frame of OO-6
MDKSSTMLVHYDKGTPAVANEIKEALEGNDVEAKVDAMKKAIMLLLNGETIPQLFI
TIIRYVLPSEDHTIQKLLLLYLELIEKTDSKGKVLPEMILICQNLRNNLQHPNEYIRGVT
LRFLCRMKETEIVEPLTPSVLQNLEHRHPFVRRNAILAIMSIYKLPHGDQLFVDAPEMI
EKVLSTEQDPSAKRNAFLMLFTCAEERAVNYLLSNVDKVSDWNESLQMVVLELIRS
VCKTKPAEKGKYIKIIISLLSATSSAVIYECAGTLVSLSSAPTAIRAAANTYCQLLLSQS
DNNVKLILLDRLYELKTLHRDIMVELIIDVLRALSSPNLDIRRKTLDIALDLITHHNINE
VVQMLKKEVVKTQSGELEKNGEYRQMLIQAIHACAVKFPEVASTVVHLLMDFLGDS
NVASALDVVVFVREIIETNPKLRVSIITRLLDTFYQIRAGKVCPCALWIIGEYCLSLSEV
ESGISTITQCLGELPFYSVSEESEPTETSKKIQPTSSAMVSSRKPVILADGTYATQSAAS
ETTFSSPTVVQGSLTSGNLRALLLTGDFFLGAVVACTLTKLVLRLEEVQSSKTEVNKT
VSQALLIMVSILQLGQSPVSPHPIDNDSYERIMLCIKLLCHRNVEMKKIWLESCRQSFV
KMISEKQLREMEELKAKTQTTHAQPDDLIDFFHLKSRKGMSQLELEDQVQDDLKRA
TGEFTKDENDANKLNRILQLTGFSDPVYAEAYVTVHHYDIALEVTVINRTKETLQNL
CLELATMGDLKLVERPQNYSLAPERSMQIKANIKVSSTETGVIFGNIVYETSNVMERN
VVVLNDIHIDIMDYISPAVCSEVAFRTMWAEFEWENKVAVNTTIQNEREFLDHIIKST
NMKCLTAPSAIAGECGFLAANLYAKSVFGEDALVNLSIEKQTDGTLSGYIRIRSKTQG
IALSLGDKITLKQKGGS

SEQ ID NO:69, Nucleotide sequence of the open reading frame of OO-8
ATGGCGAAATCTCAGATCTGGTTTGGTTTTGCGTTACTCGCGTTGCTTCTGGTTTC
AGCCGTAGCTGACGATGTGGTTGTTTTGACTGACGATAGCTTCGAAAAGGAAGTT
GGTAAAGATAAAGGAGCTCTCGTCGAGTTTTACGCTCCCTGGTGTGGTCACTGCA
AGAAACTTGCTCCAGAGTATGAAAAGCTAGGGGCAAGCTTCAAGAAGGCTAAGT
CTGTGTTGATTGCAAAGGTTGATTGTGATGAGCAAAAGAGTGTCTGTACTAAATA
TGGTGTTAGTGGATACCCAACCATTCAGTGGTTTCCTAAAGGATCTCTTGAACCT
CAAAAGTATGAGGGTCCACGCAATGCTGAAGCTTTGGCTGAATACGTGAACAAG
GAAGGAGGCACCAACGTAAAATTAGCTGCAGTTCCACAAAACGTGGTTGTTTTG
ACACCTGACAATTTCGATGAGATTGTTCTGGATCAAAACAAAGATGTCCTAGTCG

AATTTTATGCACCATGGTGTGGCCACTGCAAATCACTCGCTCCCACATACGAAAA
GGTAGCCACAGTGTTTAAACAGGAAGAAGGTGTAGTCATCGCCAATTTGGATGC
TGATGCACACAAAGCCCTTGGCGAGAAATATGGAGTGAGTGGATTCCCAACATT
GAAATTCTTCCCAAAGGACAACAAAGCTGGTCACGATTATGACGGTGGCAGGGA
TTTAGATGACTTTGTAAGCTTCATCAACGAGAAATCTGGGACCAGCAGGGACAGT
AAAGGGCAGCTTACTTCAAAGGCTGGTATAGTCGAAAGCTTAGATGCTTTGGTAA
AAGAGTTAGTTGCAGCTAGTGAAGATGAGAAGAAGGCAGTGTTGTCTCGCATAG
AAGAGGAAGCAAGTACCCTTAAGGGCTCCACCACGAGGTATGGAAAGCTTTACT
TGAAACTCGCAAAGAGCTACATAGAAAAAGGTTCAGACTATGCTAGCAAAGAAA
CGGAGAGGCTTGGACGGGTGCTTGGGAAGTCGATAAGTCCAGTGAAAGCTGATG
AACTCACTCTCAAGAGAAATATCCTAACCACGTTCGTTGCTTCTTCTTAA

SEQ ID NO:70, Deduced amino acid sequence of the open reading frame of OO-8
MAKSQIWFGFALLALLLVSAVADDVVVLTDDSFEKEVGKDKGALVEFYAPWCGHC
KKLAPEYEKLGASFKKAKSVLIAKVDCDEQKSVCTKYGVSGYPTIQWFPKGSLEPQK
YEGPRNAEALAEYVNKEGGTNVKLAAVPQNVVVLTPDNFDEIVLDQNKDVLVEFY
APWCGHCKSLAPTYEKVATVFKQEEGVVIANLDADAHKALGEKYGVSGFPTLKFFP
KDNKAGHDYDGGRDLDDFVSFINEKSGTSRDSKGQLTSKAGIVESLDALVKELVAA
SEDEKKAVLSRIEEEASTLKGSTTRYGKLYLKLAKSYIEKGSDYASKETERLGRVLGK
SISPVKADELTLKRNILTTFVASS

SEQ ID NO:71, Nucleotide sequence of the open reading frame of OO-9
ATGGCGTCGAGCGATGAGCGTCCAGGAGCGTATCCGGCACGTGACGGATCAGAG
AACTTACCTCCGGGAGATCCAAAGACGATGAAGACGGTGGTGATGGATAAAGGA
GCGGCGATGATGCAATCGTTGAAACCGATCAAACAGATGAGTCTCCATTTGTGTT
CTTTCGCTTGTTATGGTCACGATCCTAGCCGTCAGATTGAAGTCAACTTCTATGTT
CATCGACTCAACCAAGACTTTCTTCAATGTGCTGTTTACGATTGCGACTCCTCTAA
ACCCCATCTCATCGGGATCGAGTATATTGTGTCGGAGAGGTTATTTGAGAGTCTT
GATCCGGAGGAGCAAAAGCTTTGGCACTCTCATGACTATGAGATCCAAACAGGC
CTTCTAGTAACTCCAAGGGTCCCTGAGCTTGTAGCTAAGACAGAGCTTGAAAATA
TTGCCAAAACTTATGGGAAGTTTTGGTGCACTTGGCAGACCGATCGCGGGGATAA
ATTGCCACTTGGTGCACCATCACTTATGATGTCACCACAAGACGTGAATATGGGA
AAGATCAAGCCAGGGCTATTGAAGAAACGTGACGATGAGTATGGAATCTCGACG
GAATCTTTGAAGACGTCTCGAGTTGGAATTATGGGACCGGAGAAGAAAAATTCG
ATGGCTGATTATTGGGTTCATCACGGAAAAGGATTAGCGGTTGACATAATCGAA
ACTGAGATGCAGAAATTGGCTCCGTTCCCGTAA

SEQ ID NO:72, Deduced amino acid sequence of the open reading frame of OO-9
MASSDERPGAYPARDGSENLPPGDPKTMKTVVMDKGAAMMQSLKPIKQMSLHLCS
FACYGHDPSRQIEVNFYVHRLNQDFLQCAVYDCDSSKPHLIGIEYIVSERLFESLDPEE
QKLWHSHDYEIQTGLLVTPRVPELVAKTELENIAKTYGKFWCTWQTDRGDKLPLGA
PSLMMSPQDVNMGKIKPGLLKKRDDEYGISTESLKTSRVGIMGPEKKNSMADYWVH
HGKGLAVDIIETEMQKLAPFP

SEQ ID NO:73, Nucleotide sequence of the open reading frame of OO-10
ATGGCGACTCTTAAGGTTTCTGATTCTGTTCCTGCTCCTTCTGATGATGCTGAGCA
ATTGAGAACCGCTTTTGAAGGATGGGGTACGAACGAGGACTTGATCATATCAAT
CTTGGCTCACAGAAGTGCTGAACAGAGGAAAGTCATCAGGCAAGCATACCACGA
AACCTACGGCGAAGACCTTCTCAAGACTCTTGACAAGGAGCTCTCTAACGATTTC
GAGAGAGCTATCTTGTTGTGGACTCTTGAACCCGGTGAGCGTGATGCTTTATTGG

CTAATGAAGCTACAAAAAGATGGACTTCAAGCAACCAAGTTCTTATGGAAGTTG
CTTGCACAAGGACATCAACGCAGCTGCTTCACGCTAGGCAAGCTTACCATGCTCG
CTACAAGAAGTCTCTTGAAGAGGACGTTGCTCACCACACTACCGGTGACTTCAGA
AAGCTTTTGGTTTCTCTTGTTACCTCATACAGGTACGAAGGAGATGAAGTGAACA
TGACATTGGCTAAGCAAGAAGCTAAGCTGGTCCATGAGAAAATCAAGGACAAGC
ACTACAATGATGAGGATGTTATTAGAATCTTGTCCACAAGAAGCAAAGCTCAGA
TCAATGCTACTTTTAACCGTTACCAAGATGATCATGGCGAGGAAATTCTCAAGAG
TCTTGAGGAAGGAGATGATGATGACAAGTTCCTTGCACTTTTGAGGTCAACCATT
CAGTGCTTGACAAGACCAGAGCTTTACTTTGTCGATGTTCTTCGTTCAGCAATCA
ACAAAACTGGAACTGATGAAGGAGCACTCACTAGAATTGTGACCACAAGAGCTG
AGATTGACTTGAAGGTCATTGGAGAGGAGTACCAGCGCAGGAACAGCATTCCTT
TGGAGAAAGCTATTACCAAAGACACTCGTGGAGATTACGAGAAGATGCTCGTCG
CACTTCTCGGTGAAGATGATGCTTAA

SEQ ID NO:74, Deduced amino acid sequence of the open reading frame of OO-10
MATLKVSDSVPAPSDDAEQLRTAFEGWGTNEDLIISILAHRSAEQRKVIRQAYHETY
GEDLLKTLDKELSNDFERAILLWTLEPGERDALLANEATKRWTSSNQVLMEVACTR
TSTQLLHARQAYHARYKKSLEEDVAHHTTGDFRKLLVSLVTSYRYEGDEVNMTLA
KQEAKLVHEKIKDKHYNDEDVIRILSTRSKAQINATFNRYQDDHGEEILKSLEEGDD
DDKFLALLRSTIQCLTRPELYFVDVLRSAINKTGTDEGALTRIVTTRAEIDLKVIGEEY
QRRNSIPLEKAITKDTRGDYEKMLVALLGEDDA

SEQ ID NO:75, Nucleotide sequence of the open reading frame of OO-11
ATGGTGGATCTATTGAACTCGGTGATGAACCTGGTGGCGCCTCCAGCGACCATGG
TGGTGATGGCCTTTGCATGGCCATTACTGTCTTTCATTAGCTTCTCCGAACGGGCT
TACAACTCTTATTTCGCCACCGAAAATATGGAAGATAAAGTAGTTGTCATCACCG
GAGCTTCATCGGCCATTGGAGAGCAAATAGCATATGAATATGCAAAAAGAGGAG
CGAATTTGGTGTTGGTGGCGAGGAGAGAGCAGAGACTGAGAGTTGTGAGTAATA
AGGCTAAACAGATTGGAGCCAACCATGTGATCATCATCGCTGCTGATGTCATCAA
AGAAGATGACTGCCGCCGTTTTATCACCCAAGCCGTCAACTATTACGGCCGCGTG
GATCATCTAGTGAATACAGCGAGTCTTGGACACACTTTTTACTTTGAGGAAGTGA
GTGACACGACTGTGTTTCCACATTTGCTGGACATAAACTTCTGGGGGAATGTTTA
TCCGACATACGTAGCGTTGCCATACCTTCACCAGACGAATGGCCGAATAGTCGTG
AATGCATCGGTTGAAAACTGGTTGCCTCTACCACGGATGAGTCTTTATTCTGCTG
CAAAAGCAGCATTAGTCAACTTCTATGAGACGCTGCGTTTCGAGCTAAATGGAG
ACGTTGGTATAACTATCGCGACTCACGGGTGGATTGGCAGTGAGATGAGTGGAG
GAAAGTTCATGCTAGAAGAAGGTGCTGAGATGCAATGGAAGGAAGAGAGAGAA
GTACCTGCAAATGGTGGACCGCTAGAGGAATTTGCAAAGATGATTGTGGCAGGA
GCTTGTAGGGGAGATGCATATGTGAAGTTTCCAAACTGGTACGATGTCTTTCTCC
TCTATCGAGTCTTCACACCGAATGTACTGAGATGGACATTCAAGTTGTTACTGTC
TACTGAGGGTACACGTAGAAGCTCCCTTGTTGGGGTCGGGTCAGGTATGCCTGTG
GATGAATCCTCTTCACAAATGAAACTTATGCTTGAAGGAGGACCACCTCGAGTTC
CTGCAAGCCCACCTAGGTATACCGCAAGCCCACCTCATTATACCGCAAGCCCACC
ACGGTATCCTGCAAGCCCACCTCGGTATCCTGCGAGCCCACCTCGGTTTTCACAG
TTTAATATCCAAGAGTTGTAA

SEQ ID NO:76, Deduced amino acid sequence of the open reading frame of OO-11
MVDLLNSVMNLVAPPATMVVMAFAWPLLSFISFSERAYNSYFATENMEDKVVVITG
ASSAIGEQIAYEYAKRGANLVLVARREQRLRVVSNKAKQIGANHVIIIAADVIKEDDC
RRFITQAVNYYGRVDHLVNTASLGHTFYFEEVSDTTVFPHLLDINFWGNVYPTYVAL

PYLHQTNGRIVVNASVENWLPLPRMSLYSAAKAALVNFYETLRFELNGDVGITIATH
GWIGSEMSGGKFMLEEGAEMQWKEEREVPANGGPLEEFAKMIVAGACRGDAYVKF
PNWYDVFLLYRVFTPNVLRWTFKLLLSTEGTRRSSLVGVGSGMPVDESSSQMKLML
EGGPPRVPASPPRYTASPPHYTASPPRYPASPPRYPASPPRFSQFNIQEL

SEQ ID NO:77, Nucleotide sequence of the open reading frame of OO-12
ATGGCTGGAAAACTCATGCACGCTCTTCAGTACAACTCTTACGGTGGTGGCGCCG
CCGGATTAGAGCATGTTCAAGTTCCGGTTCCAACACCAAAGAGTAATGAGGTTTG
CCTGAAATTAGAAGCTACTAGTCTAAACCCTGTTGATTGGAAAATTCAGAAAGG
AATGATCCGCCCATTTCTGCCCCGCAAGTTCCCCTGCATTCCAGCTACTGATGTTG
CTGGAGAGGTCGTTGAGGTTGGATCAGGAGTAAAAAATTTTAAGGCTGGTGACA
AAGTTGTAGCGGTTCTTAGCCATCTAGGTGGAGGTGGACTTGCTGAGTTCGCTGT
TGCAACCGAGAAGCTGACTGTCAAAAGACCTCAAGAAGTGGGAGCAGCTGAAGC
AGCAGCTTTACCTGTGGCGGGTCTAACCGCTCTCCAAGCTCTTACTAATCCTGCG
GGGTTGAAGCTGGATGGTACAGGCAAGAAGGCGAACATCCTGGTCACAGCAGCA
TCTGGTGGGGTTGGTCACTATGCAGTCCAGCTGGCAAAACTTGCAAATGCTCACG
TAACCGCTACATGTGGTGCCCGGAACATAGAGTTTGTCAAATCGTTGGGAGCGG
ATGAGGTTCTCGACTACAAGACTCCCGAGGGAGCCGCCCTCAAGAGTCCGTCGG
GTAAAAAATATGACGCTGTGGTCCATTGTGCAAACGGGATTCCATTTTCGGTATT
CGAACCAAATTTGTCGGAAAACGGGAAGGTGATAGACATCACACCGGGGCCTAA
TGCAATGTGGACTTATGCGGTTAAGAAAATAACCATGTCAAAGAAGCAGTTAGT
GCCACTCTTGTTGATCCCAAAAGCTGAGAATTTGGAGTTTATGGTGAATCTAGTG
AAAGAAGGGAAAGTGAAGACAGTGATTGACTCAAAGCATCCTTTGAGCAAAGCG
GAGGATGCTTGGGCCAAAAGTATCGATGGTCATGCTACTGGGAAGATCATTGTC
GAGCCATAA

SEQ ID NO:78, Deduced amino acid sequence of the open reading frame of OO-12
MAGKLMHALQYNSYGGGAAGLEHVQVPVPTPKSNEVCLKLEATSLNPVDWKIQKG
MIRPFLPRKFPCIPATDVAGEVVEVGSGVKNFKAGDKVVAVLSHLGGGGLAEFAVA
TEKLTVKRPQEVGAAEAAALPVAGLTALQALTNPAGLKLDGTGKKANILVTAASGG
VGHYAVQLAKLANAHVTATCGARNIEFVKSLGADEVLDYKTPEGAALKSPSGKKY
DAVVHCANGIPFSVFEPNLSENGKVIDITPGPNAMWTYAVKKITMSKKQLVPLLLIPK
AENLEFMVNLVKEGKVKTVIDSKHPLSKAEDAWAKSIDGHATGKIIVEP

SEQ ID NO:79, Nucleotide sequence of the open reading frame of pp82
ATGGAAATTCCCTTAGGTCGAGATGGCGAGGGTATGCAGTCAAAGCAGTGCCCG
CGCGGCCACTGGCGTCCAGCGGAAGACGACAAGCTGCGAGAACTAGTGTCCCAG
TTTGGACCTCAAAACTGGAATCTCATAGCAGAGAAACTTCAGGGTCGATCAGGG
AAAAGCTGCAGGCTACGGTGGTTCAATCAGCTGGACCCTCGCATCAACCGGCAC
CCATTCTCGGAAGAAGAGGAAGAGCGGCTGCTTATAGCACACAAGCGCTACGGC
AACAAGTGGGCATTGATCGCGCGCCTCTTTCCGGGCCGCACAGACAACGCGGTG
AAGAATCACTGGCACGTTGTGACGGCAAGACAGTCCCGTGAACGGACACGAACT
TACGGCCGTATCAAAGGTCCGGTACATCGAAGAGGCAAGGGTAACCGTATCAAT
ACCTCCGCACTTGGAAATTACCATCACGATTCGAAGGGAGCTCTCACAGCCTGGA
TTGAGTCGAAGTATGCGACAGTCGAGCAGTCTGCGGAAGGGCTCGCTAGGTCTC
CTTGTACCGGCAGAGGCTCTCCTCCTCTACCCACCGGTTTCAGTATACCGCAGAT
TTCCGGCGGCGCCTTCCATCGACCGACAAACATGAGTACTAGTCCTCTTAGCGAT
GTGACTATCGAGTCGCCAAAGTTTAGCAACTCCGAAAATGCGCAAATAATAACC
GCGCCCGTCCTGCAAAAGCCAATGGGAGATCCCAGGTCAGTATGCTTGCCGAATT
CGACTGTTTCCGACAAGCAGCAAGTGCTGCAGAGTAATTCCATCGACGGTCAGAT

CTCCTCCGGGCTCCAGACAAGCGCAATAGTAGCGCATGATGAGAAATCGGGCGT
CATTTCAATGAATCATCAAGCACCGGATATGTCCTGTGTTGGATTGAAGTCAAAT
TTTCAGGGGAGTCTCCATCCTGGCGCTGTTAGATCTTCTTGGAATCAATCCCTTCC
CCACTGTTTTGGCCACAGTAACAAGTTGGTGGAGGAGTGCAGGAGTTCTACAGG
CGCATGCACTGAACGCTCTGAGATTCTGCAAGAACAGCATTCTAGCCTTCAGTTT
AAATGCAGCACTGCGTACAATACTGGAAGATATCAACATGAAAACCTTTGTGGG
CCAGCATTCTCGCAACAAGACACAGCGAACGAGGTTGCGAATTTTTCTACGTTGG
CATTCTCCGGCCTAGTGAAGCATCGCCAAGAGAGGTTGTGCAAAGATAGTGGAT
CTGCTCTCAAGCTGGGACTATCATGGGTTACATCCGATAGCACTCTTGACTTGAG
TGTTGCCAAAATGTCAGCATCGCAGCCAGAGCAGTCTGCGCCGGTTGCATTCATT
GATTTTCTAGGCGTGGGAGCGGCCTGA

SEQ ID NO:80, Deduced amino acid sequence of the open reading frame of pp82
MEIPLGRDGEGMQSKQCPRGHWRPAEDDKLRELVSQFGPQNWNLIAEKLQGRSGKS
CRLRWFNQLDPRINRHPFSEEEEERLLIAHKRYGNKWALIARLFPGRTDNAVKNHW
HVVTARQSRERTRTYGRIKGPVHRRGKGNRINTSALGNYHHDSKGALTAWIESKYA
TVEQSAEGLARSPCTGRGSPPLPTGFSIPQISGGAFHRPTNMSTSPLSDVTIESPKFSNS
ENAQIITAPVLQKPMGDPRSVCLPNSTVSDKQQVLQSNSIDGQISSGLQTSAIVAHDE
KSGVISMNHQAPDMSCVGLKSNFQGSLHPGAVRSSWNQSLPHCFGHSNKLVEECRS
STGACTERSEILQEQHSSLQFKCSTAYNTGRYQHENLCGPAFSQQDTANEVANFSTL
AFSGLVKHRQERLCKDSGSALKLGLSWVTSDSTLDLSVAKMSASQPEQSAPVAFIDF
LGVGAA

SEQ ID NO:81, Nucleotide sequence of the open reading frame of Pk225
ATGGAGATGAACATTAAGTTTCCAGTTATAGACTTGTCTAAGCTCAATGGTGAAG
AGAGAGACCAAACCATGGCTTTGATCGACGATGCTTGTCAAAACTGGGGCTTCTT
CGAGCTGGTGAACCATGGACTACCATATGATCTAATGGACAACATTGAGAGGAT
GACAAAGGAACACTACAAGAAACATATGGAACAAAAGTTCAAAGAAATGCTTCG
TTCCAAAGGTTTAGATACCCTCGAGACCGAAGTTGAAGATGTCGATTGGGAAAG
CACTTTCTACCTCCATCATCTCCCTCAATCTAACCTATACGACATCCCTGATATGT
CAAATGAATACCGATTGGCAATGAAGGATTTTGGGAAGAGGCTTGAGATTCTAG
CTGAAGAGCTATTGGACTTGTTGTGTGAGAATCTAGGGTTGGAGAAAGGGTACTT
GAAGAAGGTGTTTCATGGGACAACGGGTCCAACTTTTGCGACAAAGCTTAGCAA
CTATCCACCATGTCCTAAACCAGAGATGATCAAAGGGCTTAGGGCTCACACAGA
TGCAGGAGGCCTCATTTTGCTGTTTCAAGATGATAAGGTCAGTGGTCTCCAGCTT
CTTAAAGATGGTGATTGGGTTGATGTTCCTCCTCTCAAGCATTCCATTGTCATCAA
CCTTGGTGACCAACTTGAGGTGATAACAAACGGGAAGTACAAGAGTGTAATGCA
CCGTGTGATGACCCAGAAAGAAGGAAACAGGATGTCTATCGCGTCGTTTTACAA
CCCCGGAAGCGATGCTGAGATCTCTCCGGCAACATCTCTTGTGGATAAAGACTCA
AAATACCCAAGCTTTGTGTTTGATGACTACATGAAACTCTATGCCGGACTCAAGT
TTCAGGCCAAGGAGCCACGGTTCGAGGCGATGAAAAATGCTGAAGCAGCTGCGG
ATTTGAATCCGGTGGCTGTGGTTGAGACATTCTAA

SEQ ID NO:82, Deduced amino acid sequence of the open reading frame of Pk225
MEMNIKFPVIDLSKLNGEERDQTMALIDDACQNWGFFELVNHGLPYDLMDNIERMT
KEHYKKHMEQKFKEMLRSKGLDTLETEVEDVDWESTFYLHHLPQSNLYDIPDMSNE
YRLAMKDFGKRLEILAEELLDLLCENLGLEKGYLKKVFHGTTGPTFATKLSNYPPCP

KPEMIKGLRAHTDAGGLILLFQDDKVSGLQLLKDGDWVDVPPLKHSIVINLGDQLEV
ITNGKYKSVMHRVMTQKEGNRMSIASFYNPGSDAEISPATSLVDKDSKYPSFVFDDY
MKLYAGLKFQAKEPRFEAMKNAEAAADLNPVAVVETF